

(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:16:03 2000; Maspar time 9.87 Seconds

Tabular output not generated. 731.891 Million cell updates/sec

Title: >US-09-308-435-2
Description: (28-260) from US09308435.pep
Perfect Score: 1594
Sequence: 1 CSPHIETNEVALKLNHPA.....QKNLESYQKAKELGKRR 233

Scoring table: PAM 150
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 48.589; Variance 103.194; scale 0.471

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-----------------------------------|-----------|
| 1 | 1594 | 100.0 | 260 | 1 | HPA3_HELPY NEURAMINYLLACTOSE-BIND | 1.03e-279 |
| 2 | 1580 | 99.1 | 260 | 1 | HPA0_HELPY NEURAMINYLLACTOSE-BIND | 6.22e-277 |
| 3 | 1578 | 99.0 | 260 | 1 | HPA2_HELPY NEURAMINYLLACTOSE-BIND | 1.55e-276 |
| 4 | 1551 | 97.3 | 260 | 1 | HPA1_HELPY NEURAMINYLLACTOSE-BIND | 3.53e-271 |
| 5 | 848 | 53.2 | 125 | 1 | HPA4_HELPY NEURAMINYLLACTOSE-BIND | 3.78e-133 |
| 6 | 771 | 48.4 | 125 | 1 | HPA5_HELPY NEURAMINYLLACTOSE-BIND | 2.63e-118 |
| 7 | 107 | 6.7 | 616 | 1 | ORC3_YEAST ORIGIN RECOGNITION COM | 4.42e-01 |
| 8 | 105 | 6.6 | 805 | 1 | STFB_MYCPN PHENYLALANYL-TRNA SYNT | 7.61e-01 |
| 9 | 105 | 6.6 | 1803 | 1 | STFB_MYCPN TRANSPOSON TY4 207.7 K | 9.95e-01 |
| 10 | 104 | 6.5 | 313 | 1 | MEMX_NEILA CYTOSINE-SPECIFIC METH | 1.30e+00 |
| 11 | 103 | 6.5 | 542 | 1 | KPRX_YARLI PYRUVATE KINASE (EC 2. | 9.95e-01 |
| 12 | 104 | 6.5 | 616 | 1 | Y329_METJA HYPOTHETICAL PROTEIN M | 1.59e+00 |
| 13 | 102 | 6.4 | 1142 | 1 | KPE1_NEUCR PROTEIN KINASE C-LIKE | 1.59e+00 |
| 14 | 100 | 6.3 | 721 | 1 | RJ3_CHLTR 50S RIBOSOMAL PROTEIN | 2.20e+00 |
| 15 | 101 | 6.3 | 767 | 1 | YNY7_YEAST PUTATIVE CYSTEINYL-TRN | 2.86e+00 |
| 16 | 100 | 6.3 | 1251 | 1 | RBP2_PLAVB RETICULOCYTE BINDING P | 2.86e+00 |
| 17 | 98 | 6.1 | 311 | 1 | FMT_MYCPN METHIONYL-TRNA FORMYLT | 4.78e+00 |
| 18 | 98 | 6.1 | 330 | 1 | YNY7_YEAST HYPOTHETICAL 37.0 KDA | 4.78e+00 |
| 19 | 98 | 6.1 | 330 | 1 | EMPA_CLOYS ELECTRON TRANSFER FLAV | 4.78e+00 |
| 20 | 98 | 6.1 | 705 | 1 | YNY7_YEAST HYPOTHETICAL 81.5 KDA | 4.78e+00 |
| 21 | 98 | 6.1 | 764 | 1 | SYFB_HELPY PHENYLALANYL-TRNA SYNT | 4.78e+00 |
| 22 | 98 | 6.1 | 1858 | 1 | PKR2_DICDI PHOSPHATIDYLINOSITOL 3 | 1.02e+01 |
| 23 | 95 | 6.0 | 197 | 1 | RQVA_BORBU HOLLIDAY JUNCTION DNA | 1.02e+01 |

ALIGNMENTS

| RESULT | ID | HPA3_HELPY | STANDARD: | PRT: | 260 AA. |
|--------|----|------------|-----------|------|------------------------------------|
| 24 | 95 | 6.0 | 296 | 1 | COPE_YEAST |
| 25 | 96 | 6.0 | 368 | 1 | KICR_XENLA KEATININ, TYPE I CYTOSK |
| 26 | 95 | 6.0 | 404 | 1 | RPAL_SALTY O-ANTIGEN LIGASE. |
| 27 | 95 | 6.0 | 476 | 1 | YN95_YEAST POTATIVE 49.7 KDA MEMB |
| 28 | 96 | 6.0 | 671 | 1 | FADH_ECOLI 2,4-DIENOYL-COA REDUCT |
| 29 | 95 | 6.0 | 845 | 1 | SCPI_MESAU SYNAPTONEMAL COMPLEX P |
| 30 | 96 | 6.0 | 866 | 1 | LOXI_LENCU LIPIDOXENASE (EC 1.13. |
| 31 | 95 | 6.0 | 917 | 1 | VG13_YEAST HYPOTHETICAL 104.8 KDA |
| 32 | 95 | 6.0 | 1140 | 1 | XPE_CERAE POSSIBLE DNA-REPAIR PR |
| 33 | 95 | 6.0 | 1570 | 1 | PKR1_DICDI PHOSPHATIDYLINOSITOL 3 |
| 34 | 94 | 5.9 | 294 | 1 | NUCC_MOUSE ACIDIC RIBOSOMAL PROTE |
| 35 | 94 | 5.9 | 338 | 1 | RLAO_METTL ACIDIC RIBOSOMAL PROTE |
| 36 | 94 | 5.9 | 476 | 1 | PURX_WHEAT ADENYLOSUCINATE SYNTH |
| 37 | 94 | 5.9 | 501 | 1 | CIC2_HUMAN DIHYDROPIRIDINE-SENSIT |
| 38 | 94 | 5.9 | 1081 | 1 | YV02_METJA HYPOTHETICAL PROTEIN M |
| 39 | 94 | 5.9 | 1181 | 1 | SPCA_DROME SPECIRIN ALPHA CHAIN. |
| 40 | 94 | 5.9 | 2415 | 1 | HA03_YEAST HAP3 TRANSCRIPTIONAL A |
| 41 | 93 | 5.8 | 144 | 1 | PTH_HELPY PEPTIDYL-TRNA HYDROLAS |
| 42 | 93 | 5.8 | 186 | 1 | PGK_CHLPP PHOSPHOGLYCERATE KINAS |
| 43 | 93 | 5.8 | 402 | 1 | HIS2_PICPA HISTIDINE BIOSYNTHESIS |
| 44 | 93 | 5.8 | 882 | 1 | VGFL_IBVB F1 PROTEIN. |
| 45 | 93 | 5.8 | 3951 | 1 | |

SEQUENCE FROM N.A.
RC STRAIN-NCIC 11637;
RX MEDLINE: 97431527.
RA Jones A.C., Logan R.P., Foynes S., Cockayne A., Wren B.W., Penn C.W.;
RT "A flagellar sheath protein of *Helicobacter pylori* is identical to
RT HpaA, a putative N-acetylneuraminyllactose-binding hemagglutinin, but
RT is not an adhesin for AGS cells."
RL J. Bacteriol. 179:5643-5647(1997).
-1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
ANCHOR (PROBABLE).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL: X92802; CA63246.1; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR FLAGELLIN; Outer membrane; Lipoprotein; Signal.
KW SIGNAL
FT CHAIN 1 27
FT LIPID 28 260
FT DOMAIN 134 139
FT BINDING MOTIF (POTENTIAL).
SEQUENCE 260 AA; 29100 MW; F613B93FE1FF1FA0 CRC64;

Query Match 100.0%; Score 1594; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.03e-279;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 CSPIIETNEVALKINTYHPASEKVOALDEKILLRPAFOYSDNIAKEYENKKNOTALKV 87
 OY 28 CSPIIETNEVALKINTYHPASEKVOALDEKILLRPAFOYSDNIAKEYENKKNOTALKV 87
 Db 88 EOLIONOGKYIVSDSDKDFSFQAKKEGYLAAMNGEIVLRPDKRTIQKSEGLIF 147
 OY 88 EOLIONOGKYIVSDSDKDFSFQAKKEGYLAAMNGEIVLRPDKRTIQKSEGLIF 147
 Db 148 STGLDKMGVILPAGFIVTILPEPMSGESLDSFTMDLSLDIOEKFLLKTHSHSGGLVS 207
 OY 148 STGLDKMGVILPAGFIVTILPEPMSGESLDSFTMDLSLDIOEKFLLKTHSHSGGLVS 207
 Db 208 TWVKGTDNSNDIAKSAKLNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
 OY 208 TWVKGTDNSNDIAKSAKLNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
 RESULT 2
 ID HPA0_HELPY STANDARD; PRT: 260 AA.
 AC P55969;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NEURAMINYLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-ACETYLNEURAMINYLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
 GN HPA0 OR HP0797.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695 / ATCC 700392;
 RX MEDLINE: 97394467.
 RA Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., Fieischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A., Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S., Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A., McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K., Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M., Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E., Hayes W.S., Bordovsky M., Karp P.D., Smith H.O., Fraser C.M., Venter J.C.;
 RA "The complete genome sequence of the gastric pathogen Helicobacter pylori".
 RT Nature 388:539-547(1997).
 RL Nature 388:539-547(1997).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID ANCHOR (PROBABLE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC EMBL: AE000591; AAD07844.1;
 DR TIGR: HP0797;
 DR PROSITE: PS00013; PROKAR_LIPOROTIN; 1.
 KW Flagella; Outer membrane; Lipoprotein; Signal.
 FT SIGNAL 1 27 BY SIMILARITY.
 FT CHAIN 28 260 NEURAMINYLACTOSE-BINDING HEMAGGLUTININ.
 FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
 FT SEQUENCE 260 AA; 29040 MW; D9F86C8B0556D071 CRC64;
 Query Match 99.1%; Score 1580; DB 1; Length 260;
 Best Local Similarity 98.7%; Pred. No. 6 22e-277;
 Matches 230; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 88 EOLIONOGKYIVSDSDKDFSFQAKKEGYLAAMNGEIVLRPDKRTIQKSEGLIF 147
 OY 88 EOLIONOGKYIVSDSDKDFSFQAKKEGYLAAMNGEIVLRPDKRTIQKSEGLIF 147
 Db 148 STGLDKMGVILPAGFIVTILPEPMSGESLDSFTMDLSLDIOEKFLLKTHSHSGGLVS 207
 OY 148 STGLDKMGVILPAGFIVTILPEPMSGESLDSFTMDLSLDIOEKFLLKTHSHSGGLVS 207
 Db 208 TWVKGTDNSNDIAKSAKLNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
 OY 208 TWVKGTDNSNDIAKSAKLNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
 RESULT 3
 ID HPA2_HELPY STANDARD; PRT: 260 AA.
 AC Q48254;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEURAMINYLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-ACETYLNEURAMINYLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
 GN HPA0.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CCUG 17874;
 RX MEDLINE: 96042077.
 RA O'Toole P.W., Janson L., Dolg P., Huang J., Kostzynska M., Trust T.J.;
 RA "The putative neuraminylactose-binding hemagglutinin HpaA of Helicobacter pylori CCUG 17874 is a lipoprotein".
 RL J. Bacteriol. 177:6049-6057(1995).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID ANCHOR (PROBABLE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC EMBL: U35455; AAA79096.1;
 DR PROSITE: PS00013; PROKAR_LIPOROTIN; 1.
 KW Flagella; Outer membrane; Lipoprotein; Signal.
 FT SIGNAL 1 27 BY SIMILARITY.
 FT CHAIN 28 260 NEURAMINYLACTOSE-BINDING HEMAGGLUTININ.
 FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
 FT DOMAIN 134 139 N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE BINDING MOTIF (POTENTIAL).
 FT SEQUENCE 260 AA; 29003 MW; 676DDA9352E9E03C CRC64;
 Query Match 99.0%; Score 1578; DB 1; Length 260;
 Best Local Similarity 97.9%; Pred. No. 1.55e-276;
 Matches 228; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 208 TMVGTGDSNDIAIKSALNKIFPSIMQIDKILTKQKLYESQDAKELGKRR 260
 QY 208 TMVGTGDSNDIAIKSALNKIFANIMOEDIKLTKQKLYESQDAKELGKRR 260

RESULT 4
 ID HPA1_HELPY STANDARD: PRT: 260 AA.
 AC 048264;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
 DE ACTINERAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
 DE BINDING SUBUNIT) (NLBH) (FLAGELLAR SHEATH ADHESIN).
 GN HPA1.
 OS Helicobacter pylori (Campylobacter pylori).
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-8826;
 RX MEDLINE: 93139035.
 RA Evans D.G., Karjalainen T.K., Evans D.J., Graham D.Y., Lee C.-H.;
 RT "Cloning, nucleotide sequence, and expression of a gene encoding an
 RT adhesin subunit protein of Helicobacter pylori.";
 RL J. Bacteriol. 175:674-683(1993).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
 CC ANCHOR (PROBABLE).
 CC -1- PPM: THE N-TERMINUS IS BLOCKED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

DR EMBL: X61574; CAA43773.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1
 KM Flagella; Outer membrane; Lipoprotein; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 260 NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ.
 FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
 FT DOMAIN 134 139 N-ACETYLL-NEURAMINYLL-ALPHA(2,3)-LACTOSE
 FT BINDING MOTIF (POTENTIAL).
 FT SEQUENCE 260 AA: 29166 MW: 224893980657B14 CRC64;

Query Match 97.3%; Score 1551; DB 1; Length 260;
 Best Local Similarity 95.7%; Pred. No. 3, 53e-271;
 Matches 223; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

Db 28 CSPIIETNEVALKINYPASEKVOALDEKILLPAPFOYSDNIAKEYENKFNQTLKV 87
 QY 28 CSPIIETNEVALKINYPASEKVOALDEKILLPAPFOYSDNIAKEYENKFNQTLKV 87

Db 88 EELIONGKYIVNDSSDKDPSFAOKKEGYLAVMIGEIVLRPPKRTIOKSSPGLLF 147
 QY 88 EELIONGKYIVNDSSDKDPSFAOKKEGYLAVMIGEIVLRPPKRTIOKSSPGLLF 147

Db 148 STGLDKMEGYVLPAGFYVTILPEPMSGESLDSFTMDLSELDIOEFLKTTTSSHSGGLVS 207
 QY 148 STGLDKMEGYVLPAGFYVTILPEPMSGESLDSFTMDLSELDIOEFLKTTTSSHSGGLVS 207

Db 208 TMVGTGDSNDIAIKSALNKIFASIMQEDKILTKQKLYESQDAKELGKRR 260
 QY 208 TMVGTGDSNDIAIKSALNKIFANIMOEDIKLTKQKLYESQDAKELGKRR 260

RESULT 5
 ID HPA1_HELPY STANDARD: PRT: 125 AA.
 AC 048244;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLL-NEURAMINYLLACTOSE-
 DE BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
 DE (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
 GN HPA1 OR HNA1.
 OS Helicobacter nemestrinae.
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 96001251.
 RA Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
 RA Bronsdon M.A., Evans D.J., Jr.;
 RT "Genetic evidence for host specificity in the adhesin-encoding genes
 RT hxaA of Helicobacter acinonyx, hnaA of H. nemestrinae and hpaA of H.
 RT pylori.";
 RL Gene 163:97-102(1995).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
 CC ANCHOR (PROBABLE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

DR EMBL: U27127; AA85564.1;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
 KM Flagella; Outer membrane; Lipoprotein.
 FT NON_TER 1 1
 FT DOMAIN 92 97
 FT BINDING MOTIF (POTENTIAL).
 FT NON_TER 125 125
 FT SEQUENCE 125 AA: 14157 MW: 0BA29519F2B311AF CRC64;

Query Match 53.2%; Score 848; DB 1; Length 125;
 Best Local Similarity 96.0%; Pred. No. 3, 78e-133;
 Matches 120; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 1 NYHPASEKVOALDEKILLPAPFOYSDNIAKEYENKFNQTLKVEQLIONGKYIVND 60
 QY 43 NYHPASEKVOALDEKILLPAPFOYSDNIAKEYENKFNQTLKVEQLIONGKYIVND 102

Db 61 SSDKDLSEFAOKKEGYLAVMIGEIVLRPPKRTIOKSEPLSTGLDKMEGYVLPAG 120
 QY 103 SSDKDLSEFAOKKEGYLAVMIGEIVLRPPKRTIOKSEPLSTGLDKMEGYVLPAG 162

Db 121 FVKVT 125
 QY 163 FIKVT 167

RESULT 6
 ID HPA1_HELPY STANDARD: PRT: 125 AA.
 AC 047947;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLL-NEURAMINYLLACTOSE-
 DE BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
 DE (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
 GN HPA1 OR HXA1.
 OS Helicobacter acinonyx.
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 96001251.
 RA Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
 RA Bronsdon M.A., Evans D.J., Jr.;

```

RT      "Genetic evidence for host specificity in the adhesin-encoding genes
RT      hxaA of Helicobacter acinonyx, hxaA of H. nemeistrinae and hpaA of H.
RT      pylori."
RT      Gene 163:97-102(1995).
CC      -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC      ANCHOR (PROBABLE).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U27126; AAA85563.1; -
DR      PROSITE; PS0013; PROKAR_LIPOPROTEIN; PARTIAL.
DR      Flagella; Outer membrane; Lipoprotein.
FT      DOMAIN 1 97
FT      NON_TER 125 125
FT      SEQUENCE 125 AA; 13923 MW; 8060097C2E6573E2 CRC64;
Query Match 48.4%; Score 771; DB 1; Length 125;
Best Local Similarity 85.6%; Pred. No. 2.63e-118;
Matches 107; Conservative 14; Mismatches 4; Indels 0; Gaps 0;
DB      1 NYHPASETIOALDENILLKPAFOYSDNNAKYEKKCKNOILKYEELIQNGYVIVSD 60
QY      43 NYHPASEVQALDEKILLRPAFOYSDNNAKYEKKCKNOILKYEELIQNGYVIVSD 102
DB      61 SSDKDDLSFAQKKEGYLTLSGSEIVLRPDPKRTOKKSEPLSTGLDKKQVYLISAG 120
QY      103 SSDKDDLSFAQKKEGYLTLSGSEIVLRPDPKRTOKKSEPLSTGLDKKQVYLISAG 162
DB      121 FVKYV 125
QY      163 FIKYV 167
RESULT 7
ID      ORC3_YEAST STANDARD; PRT; 616 AA.
AC      P54790;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      ORIGIN RECOGNITION COMPLEX SUBUNIT 3 (ORIGIN RECOGNITION COMPLEX
DE      PROTEIN 62 KDA SUBUNIT).
GN      ORC3 OR OAF1 OR OIF1 OR YIL004M OR I1365.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC      Saccharomycetaceae; Saccharomyces.
RN      [1]
RP      SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX      MEDLINE; 96069857.
RA      Bell S.P., Mitchell J., Leber J., Kobayashi R., Stillman B.;
RT      "The multidomain structure of Orcp reveals similarity to regulators
RT      of DNA replication and transcriptional silencing."
RT      Cell 83:563-568(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 96239540.
RA      Hardy C.F.J.;
RT      "Characterization of an essential Orc2p-associated factor that plays
RT      a role in DNA replication."
RT      Mol. Cell. Biol. 16:1832-1841(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      STRAIN-S288C / FY23;
RX      MEDLINE; 96405918.
RA      Miosga T., Zimmermann F.K.;
RT      "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on

```

```

RT      a 43.7 kb fragment of chromosome XII including an open reading frame
RT      homologous to the human cystic fibrosis transmembrane conductance
RT      regulator protein CFTR."
RT      Yeast 12:693-708(1996).
CC      -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
CC      BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
CC      REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
CC      THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION.
CC      -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS OF 120 KDA, 71 KDA, 62
CC      KDA, 56 KDA, 53 KDA AND 50 KDA.
CC      -1- SUBCELLULAR LOCATION: NUCLEAR.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U34861; AAB38249.1; -
DR      EMBL; U48888; AAC49270.1; -
DR      EMBL; X91488; CAA62765.1; -
DR      EMBL; Z73109; CAA97447.1; -
DR      SGD; L0003074; ORC3.
DR      DNA replication; Nuclear protein; DNA-binding.
KW      SEQUENCE 616 AA; 72077 MW; 3FAAE3F64503F9C CRC64;
Query Match 6.7%; Score 107; DB 1; Length 616;
Best Local Similarity 25.7%; Pred. No. 4.42e-01;
Matches 26; Conservative 33; Mismatches 32; Indels 10; Gaps 10;
DB      197 KDLAVENKVDSDINENTLDFILLKSAFYDVKSLFENINTNLSEKNRQSTI 256
QY      36 NEVALKLVTHPA-SEKVALDEKILLRPAFOYSDNNAKYEKKCKNOILKYEELIQNGYVIVSD 90
DB      257 RLKKNYHKLDPV-SSNKG-FKXNGIQFQSLD-TVQGRINTL 294
QY      91 LQNGYKIVISVSDSDKDDLSFAQKKEGYLTLSGSEIVLRPDPKRTOKKSEPLSTGLDKKQVYLISAG 129
RESULT 8
ID      SYFB_MYCPN STANDARD; PRT; 805 AA.
AC      P75563;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-FEB-2000 (Rel. 39, Last annotation update)
DE      PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
DE      TRNA LIGASE BETA CHAIN) (PHERS).
GN      PHER OR MP049.
OS      Mycoplasma pneumoniae.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC      Mycoplasmataceae; Mycoplasma.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-ATCC 29342 / M129;
RX      MEDLINE; 97105885.
RA      Himmelfreisch R., Hilbert H., Plogens H., Pirkel E., Li B.-C.,
RA      Herrmann R.;
RT      "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT      pneumoniae."
RT      Nucleic Acids Res. 24:4420-4449(1996).
RT      Nucleic Acids Res. 24:4420-4449(1996).
CC      -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
CC      PYROPHOSPHATE + L-PHENYLALANINE-TRNA(PHE).
CC      -1- SUBUNIT: Tetramer of two alpha and two beta chains
CC      (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial

```

[illegible]

```
DE CYTOSINE-SPECIFIC METHYLTRANSFERASE NLAX (EC 2.1.1.73) (M.NLAX).  
OC Neisseria lactamica.  
RT Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 23970 / NRCC 2118;  
RX MEDLINE: 91117164.  
RA Labbe D., Hoeltke H.J.; Lau P.C.K.;  
RT "Cloning and characterization of two tandemly arranged DNA  
methyltransferase genes of Neisseria lactamica: an adenine-specific  
Rt M.NalII and a cytosine-type methylase.";  
RL Mol. Genet. 224:101-110(1990).  
CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
CC CCGGG. CAUSES SPECIFIC METHYLATION ON C-2 ON BOTH STRANDS.  
CC CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE =  
CC S-ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYL-CYTOSINE.  
CC -1- SIMILARITY: STRONG, TO OTHER C-5-DNA METHYLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on ways  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to licenses@isb-sib.ch).
```

```
DR EMBL: X54485; CAA38357.1; -.  
DR PIR: S12037; XYNHCL.  
DR HSSP: P05102; 5MHX.  
DR REBASE: RB02158; M.NLax.  
DR PFAM: PF00145; DNA_methylase_1.  
DR PRINTS: PR00105; CSMETRFASR.  
DR PROSITE: PS00094; C5_MTASE_1; 1.  
DR PROSITE: PS00095; C5_MTASE_2; 1.  
KW Transferase; Methyltransferase; Restriction system.  
FT ACET SITE 74 BY SIMILARTY  
SQ SEQUENCE 313 AA; 34842 MW; E7515C0C8A8BA611 CRC64;
```

```
Query Match 6.5%; Score 104; DB 1; Length 313;  
Best Local Similarity 34.9%; Pred. NO. 9,95e+01;  
Matches 29; Conservative 18; Mismatches 27; Indels 9; Gaps 9;
```

```
Dd 98 EXILLAKPPQAFLL-ENV-KQLKGHDKGR-T-LQVLAHLGQAIGKYIVTEVLKARD-FGIP 153  
|::|| |::|| |::|| |::|| |::|| |::|| |::|| |:  
Oy 56 EXILLLR-P-AFOYSDDNIAKEYEKKFKNQTALTKVEOI-LQNQGKVIVSDSSDDEFSA 112  
|::|| |::|| |::|| |::|| |::|| |::|| |:  
Db 154 QNRERIVLVGFNLGHVDYFRFPQP 176  
|::|| |::|| |::|| |::|| |::|| |::|| |:  
Cy 113 OKKES-YLAVANNGEITVLR-BDP 133
```

```
RESULT 11  
ID KPYK YARLI STANDARD; FR: 542 AA.  
AC P30614; 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DI 01-NOV-1995 (Rel. 32, Last annotation update)  
DE PYRVUATE KINASE (EC 2.7.1.40).  
OS Yarrowia lipolytica (Candida lipolytica).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
OC Dipodascaceae; Yarrowia.  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-NREL Y-1094;  
RX MEDLINE: 92380493.  
RA Strick C.A.; James L.C.; O'Donnell M.M.; Golisher M.G.; Franke A.E.;  
RT "The isolation and characterization of the pyruvate kinase-encoding  
RT gene from the yeast Yarrowia lipolytica."  
RL Gene 118:65-72(1992).  
[2]  
EP ERRATA.
```

```
RX MEDLINE: 94117073.  
RA Strick C.A.; James L.C.; O'Donnell M.M.; Golisher M.G.; Franke A.E.;
```


Query Match 6.4%; Score 102; DB 1; Length 1142;
 Best Local Similarity 23.2%; Pred. No. 1.69e+00;
 Matches 22; Conservative 28; Mismatches 39; Indels 6; Gaps 6;

DB 199 AAKRVES-KORIVLKKALRKREELHIDVDGPDSDSINIPALRKPSGLSTIRLAV 257
 QY 46 PASEVQALDEKILLRPAF-QYSD-NIAKEYNKFNKTALK-VEQILQNO-GYKIVISY 101
 DB 258 KDVDHAPLGRFARSPETFLAVRAEDIVVARKPSR 292
 QY 102 DSDKDDFS-FAQKKEGYLAVMNGEIVLRPDPKR 135

RESULT 14
 ID RL3-CHLTR STANDARD; PRT; 221 AA.
 C 084533;
 T 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE 50S RIBOSOMAL PROTEIN L3.
 GN RPLC OR RL3 OR CT528.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/WM-3/CX;
 RX MEDLINE; 99000809.
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.F., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RT Science 282:754-759(1998).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
 CC PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
 CC THE RIBOSOME (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 C
 DR EMBL; AE001323; AAC68129.1; -
 DR PFAM; PF00297; Ribosomal_L3; 1.
 DR PROSITE; PS00474; RIBOSOMAL_L3; FALSE_NEG.
 KM Ribosomal protein; rRNA-binding.
 SQ SEQUENCE 221 AA; 23514 MW; 065B8C1462353F5B CRC64;

Query Match 6.3%; Score 100; DB 1; Length 221;
 Best Local Similarity 36.7%; Pred. No. 2.86e+00;
 Matches 18; Conservative 13; Mismatches 15; Indels 3; Gaps 3;

DB 29 VISVDANVVAOLK-TASSDGIYNAVOMGADYQAPF-K-TIEKRFRALL 74
 QY 98 VISVDSKDDFSFAQKKEGYLAVMNGEIVLRPDPKRRIQKSEPL 146

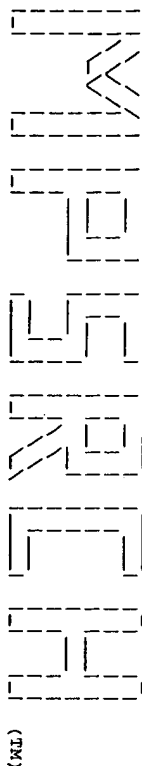
RESULT 15
 ID INT7-YEAST STANDARD; PRT; 767 AA.
 AC P53852;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C (EC 6.1.1.16) (CYSTEINE--
 DE TRNA LIGASE) (CISKS).
 GN YNL247W OR N0885.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1679;
 RX MEDLINE; 9737792.
 RA Sen-Gupta M., Gueldeuer U., Beinhauer J.D., Fiedler T.A.,
 RA Hegemann J.H.;
 RT "Sequence analysis of the 33 kb long region between ORC5 and SU11
 RT from the left arm of chromosome XIV from Saccharomyces cerevisiae,"
 RT yeast 13:849-860(1997).
 RL [2]
 RN SEQUENCE OF 44-767 FROM N.A.
 RP Poehlmann R., Philippson P.;
 RA Submitted (Mar-1996) to the EMBL/GenBank/DBJ databases.
 RL -1- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) = AMP +
 CC PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).
 CC -1- SIMILARITY: BELONGS TO CLASS-1 AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL; X96722; CAA65497.1; -
 DR EMBL; Z71523; CAA96154.1; -
 DR PFAM; PF01406; tRNA-synt_1e; 1.
 DR PRINTS; PR00983; TRNASYNTHCS.
 DR PROSITE; PS00178; AA-TRNA-LIGASE-I; FALSE_NEG.
 KM Hypothetical protein; Aminoacyl-tRNA synthetase; Protein biosynthesis;
 KM Ligase; ATP-binding.
 FT SIMILAR 65 "HIGH" REGION.
 FT 427 "KMSKS" REGION.
 FT BINDING 430 ATP (BY SIMILARITY).
 SQ SEQUENCE 767 AA; 87530 MW; 495D526781B4BE2B CRC64;

Query Match 6.3%; Score 101; DB 1; Length 767;
 Best Local Similarity 27.7%; Pred. No. 2.20e+00;
 Matches 18; Conservative 16; Mismatches 26; Indels 5; Gaps 4;

DB 101 QNVTIDDKIILRANQNYLF-DNFKENDTKFNATVVDKTKALFOYINKFTTQSEIK 159
 QY 49 EKVQALDEKILL-LRPAFGYSDNIAKEYNKFNKTALKVEQLT-Q--NOGYKIVSDSS 104
 DB 160 TIEEF 164
 QY 105 DKDDF 109

Search completed: Tue Sep 26 16:16:16 2000
 Job time : 13 secs.



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Perch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:18:57 2000; Maspar time 8.99 Seconds

Tabular output not generated. 613.770 Million cell updates/sec

Title: >US-09-308-435-4
Description: (28-260) from US09308435.pep
Perfect Score: 1599
Sequence: 1 CSPHIETNEVALKLNYPH.....QKNLESYQKAKELKGRNR 233

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 32.470; Variance 175.985; scale 0.185

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|----------|-------------------------|-----------|
| 1 | 1599 | 100.0 | 260 | 1 W07450 | Helicobacter pylori su | 2.57e-113 |
| 2 | 1599 | 100.0 | 260 | 1 W60157 | Helicobacter pylori 29 | 2.57e-113 |
| 3 | 1589 | 99.4 | 260 | 1 W60156 | Helicobacter pylori 29 | 1.58e-112 |
| 4 | 1589 | 99.4 | 260 | 1 W07449 | Helicobacter pylori su | 1.58e-112 |
| 5 | 1589 | 99.4 | 461 | 1 W74466 | Adhesin/V. cholerae tox | 1.58e-112 |
| 6 | 1547 | 96.7 | 260 | 1 Y11000 | H. pylori ORF 11ap2071 | 3.20e-109 |
| 7 | 1540 | 96.3 | 268 | 1 W20973 | H. pylori derived prot | 1.14e-108 |
| 8 | 302 | 18.9 | 97 | 1 W24663 | H. pylori derived prot | 5.74e-13 |
| 9 | 302 | 18.9 | 97 | 1 W20454 | H. pylori derived prot | 3.94e-08 |
| 10 | 235 | 14.7 | 249 | 1 W55445 | H. pylori ORF 01ae1101 | 3.94e-08 |
| 11 | 235 | 14.7 | 261 | 1 W55332 | H. pylori ORF hp3e1107 | 1.46e-07 |
| 12 | 227 | 14.2 | 249 | 1 W98332 | H. pylori GHPD 1615 pr | 5.19e-04 |
| 13 | 176 | 11.0 | 282 | 1 Y10951 | H. pylori ORF 04e1110 | 5.19e-04 |
| 14 | 176 | 11.0 | 285 | 1 W20938 | H. pylori secreted or | 3.41e+00 |
| 15 | 118 | 7.4 | 147 | 1 W20468 | H. pylori secreted or | 1.23e+01 |
| 16 | 109 | 6.8 | 549 | 1 W21638 | Grapevine leafroll vir | 1.63e+01 |
| 17 | 107 | 6.7 | 615 | 1 R77276 | ORC3 subunit of yeast | 1.63e+01 |
| 18 | 107 | 6.7 | 615 | 1 W22226 | S. cerevisiae origin o | 2.86e+01 |
| 19 | 103 | 6.4 | 2039 | 1 W56322 | Haemophilus paragailli | 2.86e+01 |
| 20 | 103 | 6.4 | 2042 | 1 W56319 | Haemophilus paragailli | 2.86e+01 |
| 21 | 100 | 6.3 | 1254 | 1 R07503 | Merozoite apical-end-10 | 4.33e+01 |
| 22 | 100 | 6.3 | 1254 | 1 W24575 | Merozoite apical-end p | 4.33e+01 |
| 23 | 99 | 6.2 | 1285 | 1 P93089 | Pasteurella multocida | 4.97e+01 |

| | | | | | | |
|----|----|-----|------|----------|-------------------------|----------|
| 24 | 97 | 6.1 | 119 | 1 W49813 | Amino acid sequence of | 6.53e+01 |
| 25 | 97 | 6.1 | 222 | 1 R39267 | Humanised C4g1 Ig heav | 6.53e+01 |
| 26 | 97 | 6.1 | 222 | 1 W49817 | Fragment of humanised | 6.53e+01 |
| 27 | 97 | 6.1 | 235 | 1 R39268 | Humanised C4g1 Ig heav | 6.53e+01 |
| 28 | 97 | 6.1 | 235 | 1 W49818 | Amino acid sequence of | 6.53e+01 |
| 29 | 97 | 6.1 | 449 | 1 R43339 | Humanised humanised C | 6.53e+01 |
| 30 | 97 | 6.1 | 449 | 1 W49816 | Human heart muscle spe | 6.53e+01 |
| 31 | 97 | 6.1 | 552 | 1 W90172 | Human heart muscle spe | 6.53e+01 |
| 32 | 97 | 6.1 | 914 | 1 W24800 | Spinocebellar ataxia | 6.53e+01 |
| 33 | 97 | 6.1 | 1312 | 1 W33807 | Human ataxin-2. | 6.53e+01 |
| 34 | 97 | 6.1 | 1313 | 1 W60213 | Spinocebellar ataxia | 6.53e+01 |
| 35 | 96 | 6.0 | 246 | 1 W95018 | Pentunia MADS box trans | 7.48e+01 |
| 36 | 96 | 6.0 | 947 | 1 R76708 | Recombinant cold-resis | 7.48e+01 |
| 37 | 96 | 6.0 | 947 | 1 R90924 | Corn derived pyruvate | 7.48e+01 |
| 38 | 95 | 5.9 | 175 | 1 R62783 | Borrelia VS461 antigen | 8.57e+01 |
| 39 | 95 | 5.9 | 599 | 1 W73479 | Grapevine leafroll vir | 8.57e+01 |
| 40 | 95 | 5.9 | 751 | 1 W13491 | Helicobacter ClpB (N-t | 8.57e+01 |
| 41 | 95 | 5.9 | 764 | 1 W98549 | H. pylori GHPD 1604 pr | 8.57e+01 |
| 42 | 94 | 5.9 | 1084 | 1 R71015 | Human neuronal calcium | 9.81e+01 |
| 43 | 94 | 5.9 | 1091 | 1 R71011 | Human neuronal calcium | 9.81e+01 |
| 44 | 95 | 5.9 | 1140 | 1 R72386 | XAR-1, part of the DNA | 9.81e+01 |
| 45 | 94 | 5.9 | 1392 | 1 Y06999 | Reslin protein sequenc | 9.81e+01 |

ALIGNMENTS

| | | |
|--------|--|---|
| RESULT | 1 | W07450 standard; Protein; 260 AA. |
| ID | W07450; | |
| AD | 04-MAR-1997 (first entry) | |
| DE | Helicobacter pylori surface-exposed 29 Kda antigen. | |
| KW | Antigen; adhesin; immunogen; vaccine; immunisation; ulcer; | |
| KM | therapy; diagnosis; | |
| OS | Helicobacter pylori strain CCUG 17874 (NCTC 11637). | |
| FH | Key | Location/Qualifiers |
| FT | peptide | 1..27 |
| FT | cleavage_site | 25..28 |
| FT | | /label= Sig_peptide |
| FT | | /note="consensus recognition site for signal |
| FT | | peptidase II in proliproteins" |
| FT | protein | 28..260 |
| FT | | /label= Mat_protein |
| FT | | /note="either the uncleaved or mature protein can |
| FT | | be used in appls. of the invention" |

| | | |
|----|---|--|
| PD | W09638475-A1. | |
| PD | 05-DEC-1996. | |
| PE | 03-JUN-1996; SE0727. | |
| PR | 01-JUN-1995; SE-002007. | |
| PR | 21-MAR-1996; SE-001085. | |
| PA | (ASTR) ASTRA AB. | |
| PI | Bolin I, Svennerholm A; | |
| DR | WPI; 97-034307/03. | |
| DR | N-PSDB; T43835. | |
| PT | Helicobacter pylori surface exposed antigen - useful for treatment, | |
| PT | propylaxis or diagnosis of mammalian infections | |
| PS | Claim 2; Page 37-38; 49pp; English. | |
| CC | A 29 kDa antigen (W07450) of Helicobacter pylori is an adhesin | |
| CC | bacterium. It is conserved in all tested strains of H. pylori, | |
| CC | and gives rise to both systemic and local (mucosal) prodn of | |
| CC | antibodies. Its amino acid sequence was deduced from a cDNA clone | |
| CC | (T43835); a sequence (W07549) deduced from a second cDNA clone | |
| CC | (T43834) differs only at position 222 (Ser for Arg). Recombinant | |
| CC | antigen can be produced in transformed host cells. It is useful | |
| CC | for the diagnosis and treatment of H. pylori infections and for | |
| CC | use as a vaccine. | |
| SO | Sequence 260 AA. | |

Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 28 CSPHIETNEVALKLNYPHASEKVALDEKILLRPAFOYSDNIAEYKFKNGTALKY 87


```

QY 28 CSPIIETNEVALKLNHPASEKVOALDEKILLRPAFYSDNIAKEENKFKNOTALKV 87
DB 88 EQLLQNGYKIVISVSDSKDQDFSAQKKEGYLAAMNGEIVLRPDPKRTIOKSEPLL 147
QY 88 EQLLQNGYKIVISVSDSKDQDFSAQKKEGYLAAMNGEIVLRPDPKRTIOKSEPLL 147
DB 148 STGLDKMEGYLIPAGFIKVTILPEMSESLDFTMDLSELDIOEKFLKTHSSHSGGLV 207
QY 148 STGLDKMEGYLIPAGFIKVTILPEMSESLDFTMDLSELDIOEKFLKTHSSHSGGLV 207
DB 208 TMYKGTDSNDIAKRALNKFANIMOEDIKLTQKNLESYQKAKELKGRNR 260
QY 208 TMYKGTDSNDIAKRALNKFANIMOEDIKLTQKNLESYQKAKELKGRNR 260

RESULT 2
ID W60157 standard; Protein; 260 AA.
AC W60157;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
N-PSDB; V34643.
M0982135-A1.
28-MAR-1998.
18-NOV-1997; SE1927.
22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
PI WPI: 98-312174/27.
DR N-PSDB; V34643.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Disclosure: Page 35-36; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34643),
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second, claimed 29 kDa antigen (see W60156) is
CC provided that differs only at residue 222 (Ser for Arg).
2 Sequence 260 AA:
Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

ID W60156 standard; Protein; 260 AA.
AC W60156;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
OS Helicobacter pylori.
PN M0982135-A1.
28-MAR-1998.
18-NOV-1997; SE1927.
22-NOV-1996; SE-004296.
PA (ASTR) ASTRA AB.
PI Berglindh T, Lofroth J;
PI WPI: 98-312174/27.
DR N-PSDB; V34642.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Claim 6; Page 28-30; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (bacillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34642),
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second 29 kDa antigen (see W60157) is provided that
CC differs only at residue 222 (Arg for Ser).
2 Sequence 260 AA:
Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 28 CSPIIETNEVALKLNHPASEKVOALDEKILLRPAFYSDNIAKEENKFKNOTALKV 87
QY 28 CSPIIETNEVALKLNHPASEKVOALDEKILLRPAFYSDNIAKEENKFKNOTALKV 87
DB 88 EQLLQNGYKIVISVSDSKDQDFSAQKKEGYLAAMNGEIVLRPDPKRTIOKSEPLL 147
QY 88 EQLLQNGYKIVISVSDSKDQDFSAQKKEGYLAAMNGEIVLRPDPKRTIOKSEPLL 147
DB 148 STGLDKMEGYLIPAGFIKVTILPEMSESLDFTMDLSELDIOEKFLKTHSSHSGGLV 207
QY 148 STGLDKMEGYLIPAGFIKVTILPEMSESLDFTMDLSELDIOEKFLKTHSSHSGGLV 207
DB 208 TMYKGTDSNDIAKRALNKFANIMOEDIKLTQKNLESYQKAKELKGRNR 260
QY 208 TMYKGTDSNDIAKRALNKFANIMOEDIKLTQKNLESYQKAKELKGRNR 260

RESULT 4
ID W07449 standard; Protein; 260 AA.
AC W07449;
DT 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW therapy; diagnosis
OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
PN Key location/Qualifiers
FT peptide 1..27
FT cleavage_site 25..28

```

FT /note= "consensus recognition site for signal
FT peptidase II in prolioproteins"
FT 28. .260
FT /label= Mat.protein
FT /note= "either the uncleaved or mature protein can
FT be used in aplns. of the invention"

PD W09638475-A1.
PD 05-DEC-1996. SE0727.
PD 03-JUN-1996; SE-002007.
PR 01-JUN-1995; SE-002007.
PR 21-MAR-1996; SE-001085.
PA (ASTR) ASTRA AB.
PI Bolin I, Svennerholm A;
DR WPI: 97-034307/03.
DR N-PSDB: T43834.
~PT Helicobacter pylori surface exposed antigen - useful for treatment,
t prophylaxis or diagnosis of mammalian infections
S Claim 2; Page 34-35; 49pp; English.
CC A 29 kDa antigen (W07449) of Helicobacter pylori is an adhesin
CC expressed on the surface of resting and dividing forms of the
CC bacterium. It is conserved in all tested strains of H. pylori,
CC and gives rise to both systemic and local (mucosal) prodn. of
CC antibodies. Its amino acid sequence was deduced from a cDNA clone
CC (T43834); a sequence (W07550) deduced from a second cDNA clone
CC (T43835) differs only at position 222 (Arg for Ser). Recombinant
CC antigen can be produced in transformed host cells. It is useful
CC for the diagnosis and treatment of H. pylori infections and for
CC use as a vaccine.
SQ Sequence 260 AA;

Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1,58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 28 CSPHIETNEVALKINYPHASEKVOALDKILLRPAFYSDINAEYENKRNQALKY 87
QY 28 CSPHIETNEVALKINYPHASEKVOALDKILLRPAFYSDINAEYENKRNQALKY 87
Db 88 E01LONOGYKVISVSDSDKDFSFQAKKEGYLVANNGEIVLRPDKRTIOKKESEGLLF 147
QY 88 E01LONOGYKVISVSDSDKDFSFQAKKEGYLVANNGEIVLRPDKRTIOKKESEGLLF 147
Db 148 STGLDKMEGVLLPAGFIKVTILEPMSGESLDSFTMDLSLDIOEKELKTHSHSGGLYS 207
148 STGLDKMEGVLLPAGFIKVTILEPMSGESLDSFTMDLSLDIOEKELKTHSHSGGLYS 207
Db 208 TWVKGTDSNDAIKSALNKIFANIMOEDIKLTOKNLESYQDAKELKGRNR 260
QY 208 TWVKGTDSNDAIKSALNKIFANIMOEDIKLTOKNLESYQDAKELKGRNR 260

RESULT 5
ID W74466 standard; Protein; 461 AA.
AC W74466;
DT 18-MAY-1999 (first entry)
DE Adhesin/V.cholerae toxin A2 and B subunit fusion protein.
KW Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW duodenal ulcer.
OS Chimeric - Helicobacter pylori.
OS Chimeric - Vibrio cholerae.
PN W09853082-A1.
PD 26-NOV-1998.
PD 21-MAY-1997; 000091.
PR 21-MAY-1997; WO-KR0091.
PA (DAEW-) DAEMOONG PHARM CO LTD.
PI Kim BO, Lee BK, Park SK, Yoon SW, Yu YH;
DR WPI: 98-434925/37.
DR N-PSDB: X21955.
PT Fusion gene consisting of the Helicobacter pylori adhesin gene
PT ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
PT useful in the diagnosis and treatment of H. pylori
PS Claim 2; Page 20-21; 32pp; English.

CC This sequence represents the fusion protein of the invention, between the
CC H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC This protein is used in the diagnosis of and as a vaccine against
CC H. pylori, which is the main cause of gastritis and gastric/duodenal
CC ulcers. It may also be used in the production of an anti-H. pylori
CC antibody. Vectors containing the DNA sequence and E. coli strains
CC containing it form the basis of a process to produce the protein, which
CC consists of culturing the microorganism and obtaining the protein. The
CC protein is an effective vaccine against H. pylori due to its excellent
CC immunogenicity for H. pylori, stability within the stomach environment,
CC and its ability to penetrate through the stomach mucous membrane to
CC stimulate sign production.
CC Note: This sequence was indexed from W09853082, which is the first
CC major country equivalent to KR97059278.
SQ Sequence 461 AA;

Query Match 99.4%; Score 1589; DB 1; Length 461;
Best Local Similarity 99.6%; Pred. No. 1,58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 39 CSPHIETNEVALKINYPHASEKVOALDKILLRPAFYSDINAEYENKRNQALKY 98
QY 28 CSPHIETNEVALKINYPHASEKVOALDKILLRPAFYSDINAEYENKRNQALKY 87
Db 99 E01LONOGYKVISVSDSDKDFSFQAKKEGYLVANNGEIVLRPDKRTIOKKESEGLLF 158
QY 88 E01LONOGYKVISVSDSDKDFSFQAKKEGYLVANNGEIVLRPDKRTIOKKESEGLLF 147
Db 159 STGLDKMEGVLLPAGFIKVTILEPMSGESLDSFTMDLSLDIOEKELKTHSHSGGLYS 218
QY 148 STGLDKMEGVLLPAGFIKVTILEPMSGESLDSFTMDLSLDIOEKELKTHSHSGGLYS 207
Db 219 TWVKGTDSNDAIKSALNKIFANIMOEDIKLTOKNLESYQDAKELKGRNR 271
QY 208 TWVKGTDSNDAIKSALNKIFANIMOEDIKLTOKNLESYQDAKELKGRNR 260

RESULT 6
ID Y11000 standard; Protein; 260 AA.
AC Y11000;
DT 08-JUN-1999 (first entry)
DE H. pylori ORF 11ap20714_4797137_f3_45 cell envelope protein.
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW secreted protein; cellular protein.
OS Helicobacter pylori.
PN W09618323-A1.
PD 07-MAY-1998.
PD 28-OCT-1997; U19575.
PR 14-JUL-1997; US-891928.
PR 28-OCT-1996; US-739150.
PR 06-DEC-1996; US-759739.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI: 98-271811/24.
DR N-PSDB: X30467.
PT Helicobacter pylori nucleic acids and proteins - used to develop
PT products for the detection, prevention and treatment of H. pylori
PT infections
PS Claims 27, 31; Page 207-208; 279pp; English.
CC Recombinant or substantially pure preparations of H. pylori polypeptides
CC are disclosed, together with the nucleic acids encoding them. In all,
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
CC secreted proteins or other cellular proteins. Vaccines containing at least
CC 8 nucleotides from the nucleic acids are claimed, as are probes containing at least
CC for treating or reducing the risk of H. pylori infections, and the
CC probes can be used diagnostically for detecting the presence of
CC Helicobacter in a sample. The products are also of use in screening
CC for compounds having the ability to interfere with the H. pylori life
CC cycle or to inhibit H. pylori infection.
SQ Sequence 260 AA;

Query Match 96.7%; Score 1547; DB 1; Length 260;


```

FH Key Location/Qualifiers
FT misc_difference 94
FT /label= Unknown
FT /note= "encoded by GAK"
PN MO9640893-A1.
PD 19-DEC-1996.
PR 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PI (ASTR ) ASTRA AB.
PI Berglundh OT. Smith D, Meligaard BL.
DR WPI: 97-052306/05.
DR N-PSDB: T67801.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
infection, and to detect Helicobacter
disclosure: Page 625; 1481pp; English.
SQ
1 The present sequence is a H. pylori derived protein, no further details
are given in the specification.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori life cycle activators or inhibitors.
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
overlapping contigs generated by mechanically shearing the bacterial
DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ
Sequence 97 AA:

Query Match 18.9%; Score 302; DB 1; Length 97;
Best Local Similarity 97.7%; Pred. No. 5,74e-13;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 53 CSPHIETNEVALKLNHPASEKVOALDEKILLRPAFOYSXNI 96
QY 28 CSPHIETNEVALKLNHPASEKVOALDEKILLRPAFOYSXNI 71

RESULT 10
W55445 standard; Protein; 249 AA.
DE 24-JUN-1998 (first entry)
DE H. pylori ORF 01ae11010_40688_c2_38 cell envelope OMP.
KM Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM identification; binding compound; bacteria; life cycle; activator;
KM inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN W09737044-A1.
PD 09-OCT-1997.
PR 27-MAR-1997; U052223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR ) ASTRA AB.
PI Alm RA, Smith D.
PI WPI: 97-503122/46.
DR N-PSDB: V24854.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
infection and for diagnosis of H. pylori infection
PS Claims 14,80; Page 653; 1145pp; English.
CC This sequence is a H. pylori cell envelope outer membrane protein (OMP).
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of

```

```

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ
Sequence 249 AA:

Query Match 14.7%; Score 235; DB 1; Length 249;
Best Local Similarity 25.1%; Pred. No. 3,94e-08;
Matches 57; Conservative 75; Mismatches 73; Indels 22; Gaps 21;

DB 40 VELHFRY-PIKGOBKNNHVLVDPRTEANKVTPENQKEPKSLFQLSNFERKGY 98
QY 38 VALKLNHPASEKVOALDEK-ILLRPAFOYSXNI AKETENFKNOTALKEQILQNOGI 96
DB 99 SY-S-QFKDVESEIPODIEKALLVLRMGQNAI-LE-D-IVEE-SDA--L-S-E-EKVID 147
QY 97 KVISVDSKQDFSPAKKEGYLAAMGEILVRDPKRTIQKSPGLFSGIDKMG 156
DB 148 -N-SSGYLNFVEPKSEDIHSFGIDVSKIAVIER-VEL-RTNNGGFVPTVHRK 203
QY 157 VLPAGFIKVTILEPMSGESLDSFTMDISEL-DIOEKFLKTHSHSGGLVS-TWV-KGT 213
DB 204 EFDHDAIKKINQAVHKMAHITKELSKHMERKVSSEAK-RKX 249
QY 214 DNSND-AIKRALNKIFANIMOEDIKTKQKLESTQKAKELDKRN 259

RESULT 11
W55332 standard; Protein; 261 AA.
AC W55332.
DE 15-JUN-1998 (first entry)
DE H. pylori ORF hpjell1075orf3 protein.
KM Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM identification; binding compound; bacteria; life cycle; activator;
KM inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT misc_difference 261
FT /note= "the nucleotides encoding this amino acid
are not given in the specification"
PN W09737044-A1.
PD 09-OCT-1997.
PR 27-MAR-1997; U052223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR ) ASTRA AB.
PI Alm RA, Smith D.
PI WPI: 97-503122/46.
DR N-PSDB: V24741.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
infection and for diagnosis of H. pylori infection
PS Claim 14; Pages 558-559; 1145pp; English.
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

```

RESULT 12
 ID W98332 standard; Protein: 249 AA.
 AC W98332.
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPD 1615 protein.
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KM peptic ulcer disease.
 NC Helicobacter pylori.
 PN W09843478-AA.
 PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881327.
 TM (HUMA-) HUMAN GENOME SCI INC.
 TI (IMR-) MERIEUX ORAYAX PASTEUR MERIEUX SERUMS.
 PI AL-GARAWI A, KLEANTHOUS H, MILLER C, OOMEN RP, TOMB J;
 DR WPI: 98-542293/46.
 DR N-PSDB: X14051.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8; Page 455-456; 2034pp; Eng1181.
 CC This sequence represents a Helicobacter pylori GHPD protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SQ Sequence 249 AA;

| | | | | |
|-----------------------|--------|---------------------|------------|----------------|
| Query Match | 14.28; | Score 227; | DB 1; | Length 249; |
| Best Local Similarity | 25.08; | Pred. NO. 1.46e-07; | | |
| Matches | 58; | Conservative | 75; | Mismatches 76; |
| | | | Indels 23; | Gaps 22; |

RESULT 13
 ID Y10951 standard; Protein; 282 AA.
 AC Y10951;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF 04ee1108_3906963.fl.7 cell envelope protein.
 KW vaccine; probe; diagnostic; ORF; cell envelope protein;
 KM secreted protein; cellular protein.
 OS Helicobacter pylori.
 PN W09818323-A1.
 PD 07-MAY-1998.
 PE 28-OCR-1997; U19575.
 PR 14-JUL-1997; US-891928.
 PR 28-OCR-1996; US-739150.
 PR 06-DEC-1996; US-755739.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI; 98-271811/24.
 DR N-PSDB; X30418.
 PT Helicobacter pylori nucleic acids and proteins - used to develop
 PT products for the detection, prevention and treatment of H. pylori
 PT infections
 PS Claims 2', 31; Page 151; 279pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
 CC secreted proteins or other cellular proteins. Vaccines containing the
 CC nucleic acids or proteins are claimed, as are probes containing at least
 CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful
 CC for treating or reducing the risk of H. pylori infections, and the
 CC probes can be used diagnostically for detecting the presence of
 CC Helicobacter in a sample. The products are also of use in screening
 CC for compounds having the ability to interfere with the H. pylori life
 CC cycle or to inhibit H. pylori infection.
 SQ Sequence 282 AA;

| | | | | |
|-----------------------|--------|---------------------|------------|----------------|
| Query Match | 11.0%; | Score 176; | DB 1; | Length 282; |
| Best Local Similarity | 25.1%; | Pred. No. 5.19e-04; | | |
| Matches | 56; | Conservative | 59; | Mismatches 90; |
| | | | Indels 18; | Gaps 14; |

| | | | |
|----|-----|--|-----|
| Db | 64 | PLBNTY-PV-HIYQAFONHHVAGTMRIVOSDUL-KPYLIDKODALINLOITFEERGY | 120 |
| QY | 39 | AKRNTYPPASEKVOA-IDEKIL-ULRPAFOYSDINAKETENKREKNOTALXVEOLLQNGY | 96 |
| Db | 121 | QVLEFOD-EKA-LNVODKKIFSVLIDKGWGILIEDLKNUL-K-DPN--SPNLDL- | 170 |
| QY | 97 | KVLSVDSDDDDDSFAQKKEGYLAVANNGEILVRPDEKRTIOKSEBGLFSTGLDMEG | 150 |
| Db | 171 | VDOSSGSVWNEFEPEBSNRVYHDFAVBVGFOALTITTYTSTNNAAGGFNSKSVIHNELD | 233 |
| QY | 157 | VLLPAGIKVTLIEPMSGSLDSFTMDLSELD-IOEKFLTHTSSHSGGGLVSTWYGT-D | 214 |
| Db | 231 | KNREDAIHKILINRKYAVVMKAVTETLKENIAYYRDAIDBMKG | 273 |
| QY | 215 | NSYDAIKRALNKIFANIMOEDIKKULQKULBESYQDAKALKG | 256 |

(TM)

Distribution rights by Oxford Molecular Ltd

generated

1 CSPH IETNEVALKLNHPA.....QKNLESYQDAKELGKRN 233

2PM 150

35661 seqs, 30989116 residues

existing first 45 summarles

1:swlssprot

mean 48.644; variance 103.561; scale 0.470

by analysis of the total score distribution.

SUMMARIES

| Accession | Protein | Length | DB | ID | Description | Pred. No. |
|-----------|---------|--------|------------|-------------------------|-------------|-----------|
| U00001 | 1858 | 1 | HPA3_HELFP | NEURAMINYLACTOSE-BIND | 5_316-278 | |
| U00001 | 773 | 1 | HPA0_HELFP | NEURAMINYLACTOSE-BIND | 3_146-275 | |
| U00001 | 315 | 260 | HPA2_HELFP | NEURAMINYLACTOSE-BIND | 7_828-275 | |
| U00001 | 337 | 260 | HPA1_HELFP | NEURAMINYLACTOSE-BIND | 1_736-269 | |
| U00001 | 312 | 125 | HPA4_HELCE | NEURAMINYLACTOSE-BIND | 8_646-1133 | |
| U00001 | 380 | 125 | HPA5_HELCE | NEURAMINYLACTOSE-BIND | 5_546-1138 | |
| U00001 | 358 | 616 | Y329_METUA | HYPOTHETICAL PROTEIN M | 2_688-01 | |
| U00001 | 378 | 616 | ORC3_YEAST | ORIGIN RECOGNITION COM | 4_636-01 | |
| U00001 | 377 | 313 | MTXA_NELIA | CYTOSINE-SPECIFIC METH | 1_046+00 | |
| U00001 | 335 | 805 | SYEP_MYCPN | PHENANTHRYL--TRNA SYNT | 1_358+00 | |
| U00001 | 334 | 1142 | KPCI_NUCDC | PROTEIN KINASE C-LIKE | 1_766+00 | |
| U00001 | 353 | 721 | RL3_YEAST | SOS RIBOSOMAL PROTEIN | 2_297+00 | |
| U00001 | 353 | 767 | Y377_CHLTR | PUTATIVE CYSTEINYL--TRN | 2_396+00 | |
| U00001 | 363 | 1251 | BBP2_YEAST | RETICULOCTE BINDING P | 2_937+00 | |
| U00001 | 363 | 1503 | Y3J3_YEAST | TRANSPOSON TY4 207.7 K | 2_976+00 | |
| U00001 | 333 | 542 | KPKR_YARLI | PRIVATIVE KINASE (EC 2. | 3_856+00 | |
| U00001 | 332 | 866 | LO1_LENCU | LIPXYGENASE (BC 1.13. | 3_856+00 | |
| U00001 | 351 | 311 | FMY_MYCPN | METHIONYL--TRNA FORNYL | 4_978+00 | |
| U00001 | 351 | 365 | RNP_YEAST | RNA POLYMERASE I SPECI | 6_408+00 | |
| U00001 | 351 | 491 | PBP_BASCU | PUTATIVE PENCILLIIN BI | 6_408+00 | |
| U00001 | 331 | 730 | EP2_MEYMT | ELONGATION FACTOR 2 (E | 6_408+00 | |
| U00001 | 331 | 773 | MAVS_YEAST | ATP-DEPENDENT RNA HELI | 6_406+00 | |
| U00001 | 1858 | 1 | P3K2_DICID | PHOSPHATIDYLINOSITOL 3 | 4_976+00 | |

| | | | | | | | |
|----|----|-----|------|---|------------|--------------------------|----------|
| 24 | 96 | 6.0 | 330 | 1 | YU7_YEAST | HYPOTHETICAL 37.0 KDA | 8.24e+00 |
| 25 | 96 | 6.0 | 330 | 1 | EFPA_CLOTS | ELECTRON TRANSFER FLAV | 8.24e+00 |
| 26 | 96 | 6.0 | 671 | 1 | PADH_ECOLI | 2,4-DIENOYL-COA REDUCT | 8.24e+00 |
| 27 | 96 | 6.0 | 705 | 1 | YU7_YEAST | HYPOTHETICAL 81.5 KDA | 8.24e+00 |
| 28 | 96 | 6.0 | 845 | 1 | SCPI_MESAU | SYNAPTONEMAL COMPLEX P | 8.24e+00 |
| 29 | 96 | 6.0 | 947 | 1 | PODM_MAZAE | PRIVATE, PHOSPHATE DIK | 8.24e+00 |
| 30 | 96 | 5.9 | 197 | 1 | RUVA_BOBBO | HOLIDAY JUNCTION DNA | 1.06e+01 |
| 31 | 94 | 5.9 | 294 | 1 | NUCG_MOUSE | ENDONUCLEASE G PRECURS | 1.35e+01 |
| 32 | 95 | 5.9 | 296 | 1 | COPE_YEAST | COTOMER EPSILON SUBUN | 1.06e+01 |
| 33 | 95 | 5.9 | 404 | 1 | RFAL_SALTY | O-ANTIGEN LIGASE. | 1.06e+01 |
| 34 | 95 | 5.9 | 436 | 1 | PUR5_YEAST | POTATIVE 49.7 KDA MEMB | 1.06e+01 |
| 35 | 94 | 5.9 | 476 | 1 | YU7_YEAST | ADENYLOSUCINATE SYNTH | 1.35e+01 |
| 36 | 94 | 5.9 | 501 | 1 | TACY_PAPAL | ALVEOLIN PRECURSOR (| 1.35e+01 |
| 37 | 95 | 5.9 | 764 | 1 | SYEP_HERBY | PHENYLANILNYL-TENRA SYTN | 1.06e+01 |
| 38 | 95 | 5.9 | 917 | 1 | YGJ3_YEAST | HYPOTHETICAL 104.8 KDA | 1.06e+01 |
| 39 | 94 | 5.9 | 1091 | 1 | CI22_HUMAN | DIDHYDROPIRIMIDIN-SENSIT | 1.35e+01 |
| 40 | 95 | 5.9 | 1140 | 1 | XPE_CEREA | POSSIBLE DNA-REPAIR PR | 1.06e+01 |
| 41 | 94 | 5.9 | 1181 | 1 | YU2_MESUA | HYPOTHETICAL PROTEIN M | 1.35e+01 |
| 42 | 94 | 5.9 | 1427 | 1 | REST_HUMAN | RESTIN (CYTOPLASMIC LI | 1.35e+01 |
| 43 | 95 | 5.9 | 1570 | 1 | PKL1_DICDI | PHOSPHATIDYLINOSITOL 3 | 1.06e+01 |
| 44 | 94 | 5.9 | 1679 | 1 | YIO9_YEAST | HYPOTHETICAL 195.1 KDA | 1.35e+01 |
| 45 | 93 | 5.8 | 845 | 1 | CC47_YEAST | CELL DIVISION CONTROL | 1.73e+01 |

ALIGNMENTS

| | | |
|---------------------------|---|---|
| ID | RESULT | 1 |
| AC | HPA3_HELPY | STANDARD; PRT; 260 AA. |
| NC | Q48261; | |
| DT | 01-NOV-1997 | (Rel. 35, Created) |
| DI | 01-NOV-1997 | (Rel. 35, Last sequence update) |
| DT | 15-JUL-1998 | (Rel. 36, Last annotation update) |
| DE | NEURAMINYLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N- | |
| DE | ACYL) NEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR- | |
| GN | BINDING SUBUNIT) (NIH) (FLAGELLAR SHEATH ADHESIN). | |
| OS | HPA3. | |
| OC | Helicobacter pylori (Campylobacter pylori). | |
| CC | Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; | |
| RC | Helicobacter. | |
| RX | [1] | |
| RZ | SEQUENCE FROM N.A. | |
| RA | STRAIN-NCTC 11637; | |
| RJ | MEDLINE; 97431527. | |
| RT | Jones A.C., Logan R.P., Foynes S., Cockayne A., Wren B.W., Penn C.W.; | |
| RL | "A flagellar sheath protein of Helicobacter pylori is identical to | |
| CC | HpaA, a putative N-acetylnauraminyllactose-binding hemagglutinin, but | |
| CC | is not an adhesin for AGS cells." ; | |
| CC | J. Bacteriol. 179:5643-5647(1997). ; | |
| CC | -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID | |
| CC | ANCHOR (PROBABLE). | |
| CC | ----- | |
| CC | This SWISS-Prot entry is copyright. It is produced through a collaboration | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | |
| CC | the European Bioinformatics Institute. There are no restrictions on ways | |
| CC | use by non-profit institutions as long as its content is in no way com- | |
| CC | mmercialized and this statement is not removed. Usage by and for commercial | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | |
| CC | or send an email to license@lsb.sib.ch). | |
| CC | ----- | |
| DR | EMBL; X92502; CA63246.1; " | |
| DR | PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1. | |
| KM | Flagella; Outer membrane; Lipoprotein; Signal. | |
| FT | SIGNAL | 1 27 |
| FT | CHAIN | 28 260 |
| FT | LIPID | 28 28 |
| FT | DOMAIN | 134 139 |
| FT | SEQUENCE | 260 AA; 29100 MW; F613B93FEF1E1F40 CRC64; |
| Query March | 99.4%; | Score 1589; DB 1; Length 260; |
| Best Local Similarity | 99.6%; | Pred. No. 5,31e 276; |
| Matches 332; Conservative | 0; | Mismatches 1; Indels 0; Gaps 0; |

```

Db 28 CSPIHITETNEVALKLNHPASEKVOALDEKILLRPAFOYSDNTAKYEENKFNQOTALKV 87
| 88 EQILQNGYKVIYVSDSDKDDFSFAOKKEGYLAAMNGEIVLRDPKRTIOKSEPGILF 147
| 148 STGLDKMEGYLIPAGFIKVTILEPMSGESLDSFTMDLSELDIOEKFLKTTSSHSGGLVS 207
| 148 STGLDKMEGYLIPAGFIKVTILEPMSGESLDSFTMDLSELDIOEKFLKTTSSHSGGLVS 207
QY 208 TMYKGTDSNDATKSAINKIPANIMOETDKLTOKNLESYOKDAKELGKRNK 260
| 208 TMYKGTDSNDATKSAINKIPANIMOETDKLTOKNLESYOKDAKELGKRNK 260
QY 208 TMYKGTDSNDATKSAINKIPANIMOETDKLTOKNLESYOKDAKELGKRNK 260

RESULT 2
ID HPA0_HELPY STANDARD; PRT; 260 AA.
AC P53969;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
BINDING SUBUNIT) (NMB) (FLAGELLAR SHEATH ADHESIN).
GN HPA0 OR HP0797.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE: 97394467.
RA Tomblin J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khakhria H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Cotton R.G., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori."
RL Nature 388:539-547(1997).
CC - SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
ANCHOR (PROBABLE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000591; AAD07844.1;
CC
DR TIGR: HP0797;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Flagella; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 27
FT CHAIN 28 260
FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 260 AA; 29040 MM; D9F6C8B0536D071 CRC64;
SQ
Query Match 98.5%; Score 1575; DB 1; Length 260;
Best Local Similarity 98.3%; Pred. No. 3,14e-275;
Matches 229; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 88 EQILQNGYKVIYVSDSDKDDFSFAOKKEGYLAAMNGEIVLRDPKRTIOKSEPGILF 147
| 88 EQILQNGYKVIYVSDSDKDDFSFAOKKEGYLAAMNGEIVLRDPKRTIOKSEPGILF 147
| 148 STGLDKMEGYLIPAGFIKVTILEPMSGESLDSFTMDLSELDIOEKFLKTTSSHSGGLVS 207
| 148 STGLDKMEGYLIPAGFIKVTILEPMSGESLDSFTMDLSELDIOEKFLKTTSSHSGGLVS 207
QY 208 TMYKGTDSNDATKSAINKIPANIMOETDKLTOKNLESYOKDAKELGKRNK 260
| 208 TMYKGTDSNDATKSAINKIPANIMOETDKLTOKNLESYOKDAKELGKRNK 260
QY 208 TMYKGTDSNDATKSAINKIPANIMOETDKLTOKNLESYOKDAKELGKRNK 260

RESULT 3
ID HPA2_HELPY STANDARD; PRT; 260 AA.
AC O48254;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-
ACETYLNEURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-
BINDING SUBUNIT) (NMB) (FLAGELLAR SHEATH ADHESIN).
GN HPA2.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CCUG 17874;
RX MEDLINE: 96042077.
RA O'Toole P.W., Janson L., Dolg P., Huang J., Kostorzynska M.,
RA Trust T.J.;
RT "The putative neuraminylactose-binding hemagglutinin Hpa2 of
RT Helicobacter pylori CCUG 17874 is a lipoprotein."
RL J. Bacteriol. 177:6049-6057(1995).
CC - SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
ANCHOR (PROBABLE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U35455; AAA79096.1;
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Flagella; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 27
FT CHAIN 28 260
FT LIPID 28 28 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 134 139 BINDING MOTIF (POTENTIAL).
FT SEQUENCE 260 AA; 29003 MM; 676DDA952E99E03C CRC64;
SQ
Query Match 98.4%; Score 1573; DB 1; Length 260;
Best Local Similarity 97.4%; Pred. No. 7.82e-275;
Matches 227; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```


| DB | 208 | THWKGTDNSNDIAIKSALNKIFGSIOMEIDDKLTQKULESOGXDAKELKGRNR | 260 |
|--------|--|--|---|
| QY | 208 | THWKGTDNSNDIAIKRALNKIFANIMQEIIDKLTQKULESIOKDAKELKGRNR | 260 |
| DB | 208 | THWKGTDNSNDIAIKSALNKIFGSIOMEIDDKLTQKULESOGXDAKELKGRNR | 260 |
| AC | 048264 | STANDARD | PRT: 260 AA. |
| DT | 01-NOV-1997 | (Rel. 35, Created) | |
| DT | 01-NOV-1997 | (Rel. 35, Last sequence update) | |
| DT | 01-NOV-1997 | (Rel. 35, Last annotation update) | |
| DE | NEURAMINYLACTOSE-BINDING HEMAGGLUTININ PRECURSOR (N-ACETYLMURAMINYLLACTOSE-BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NMBH) (FLAGELLAR SHEATH ADHESIN). | | |
| GN | HPAA. | | |
| OS | Helicobacter pylori (Campylobacter pylori). | | |
| OC | Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter. | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN-8826; | | |
| RX | MEDLINE; 93139035. | | |
| RA | Evans D.G., Karjalainen T.R., Evans D.J., Graham D.Y., Lee C.-H., | | |
| RT | "Cloning, nucleotide sequence, and expression of a gene encoding an adhesin subunit protein of Helicobacter pylori." | | |
| RL | J. Bacteriol. 175:674-683(1993). | | |
| CC | -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID ANCHOR (PROBABLE). | | |
| CC | -1- PTM: THE N-TERMINUS IS BLOCKED. | | |
| CC | ----- | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch). | | |
| CC | ----- | | |
| DR | EMBL; X61574; CAA43773.1; " | | |
| KW | PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1. | | |
| RW | Flagella; Outer membrane; Lipoprotein; Signal. | | |
| FT | SIGNAL | 1 27 | BY SIMILARITY. |
| FT | CHAIN | 28 260 | NEURAMINYLACTOSE-BINDING HEMAGGLUTININ. |
| FT | LIPID | 28 28 | N-ACYL DIGLYCERIDE (PROBABLE). |
| FT | DOMAIN | 134 139 | N-ACETYLMURAMINYL-ALPHA(2,3)-LACTOSE BINDING MOTIF (POTENTIAL). |
| FT | SEQUENCE | 260 AA; 29166 MW; 22489598065E7B14 CRC64; | |
| QY | Query Match | 96.7%; Score 1546; DB 1; Length 260; | |
| QY | Best Local Similarity | 95.3%; Pred. No. 1,73e-269; | |
| QY | Matches 222; Conservative | 8; Mismatches 3; Indels 0; Gaps 0; | |
| DB | 28 | CSPHIETNEVALKITNYHPASEKQVALDRIILKLPAPYSQNIKAKEYENKRNQTLTKV | 87 |
| QY | 28 | CSPHIETNEVALKITNYHPASEKQVALDRIILKLPAPYSQNIKAKEYENKRNQTLTKV | 87 |
| DB | 88 | EEILNOCQKIVNVSDSDDESFQAKKQKGYLAANAIGIYLRPPDKRTIOKKSEPGILF | 147 |
| QY | 88 | EEILNOCQKIVNVSDSDDESFQAKKQKGYLAANAIGIYLRPPDKRTIOKKSEPGILF | 147 |
| DB | 148 | STGLDMKESVLLIPAGFVKVYTLLEPMSSGESLDFMDLSLDIOEKFLKTTSHSHSGGLVS | 207 |
| QY | 148 | STGLDMKESVLLIPAGFVKVYTLLEPMSSGESLDFMDLSLDIOEKFLKTTSHSHSGGLVS | 207 |
| DB | 208 | THWKGTDNSNDIAIKSALNKIFASIQGEMDKLTQKULESIOKDAKELKGRNR | 260 |
| QY | 208 | THWKGTDNSNDIAIKRALNKIFANIMQEIIDKLTQKULESIOKDAKELKGRNR | 260 |
| RESULT | 5 | | |
| ID | HPAA.HELNE | STANDARD. | PRT: 125 AA. |
| AC | 048244. | | |
| DT | 01-NOV-1997 | (Rel. 35, Created) | |

```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEURAMINYLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLNURAMINYLACTOSE-
DE BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
DE (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
DE HPAa OR HNAa.
OS Helicobacter nesterinae.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96001251.
RA Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
RA Brionsdon M.A., Evans D.J. Jr.,
RT "Genetic evidence for host specificity in the adhesin-encoding genes
RT hxaA of Helicobacter acinonyx, hnaA of H. nesterinae and hpaA of H.
RT pylori."
RL Gene 163:97-102(1995).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U27127; AAA85564.1; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
KW Flagella; Outer membrane; Lipoprotein.
FT FT 1
FT DONAIN 92 97 N-ACETYL-NEURAMINYLA-ALPHA(2,3)-LACTOSE
FT BINDING MOTIF (POTENTIAL).
FT FT
FT SEQUENCE 125 AA; 14157 MW; 0BA29519F2B311AF CRC64;
SQ
Db 1 NYHPASEVYQALDEKILLRPAPFYRDNIAKEYENKFNQTLKVEQILQNGKKVINVD 60
Qy 43 NYHPASEVYQALDEKILLRPAPFYSDNIAKEYENKFNQALAKVEQLIQNGKKVISVD 102
Db 61 SSDKDDLSFAKKKSGYLAVANNGEIVLRPDPKRTIQKSESGLLFSTGLDKMEGYLIPAG 120
Qy 103 SSDKDDLSFAKKKSGYLAVANNGEIVLRPDPKRTIQKSESGLLFSTGLDKMEGYLIPAG 162
Db 121 FVKVY 125
Qy 163 FIKVT 167
RESULT 6
ID HPAa HELAC STANDARD; PRT; 125 AA.
AC 0479A7.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEURAMINYLACTOSE-BINDING HEMAGGLUTININ (N-ACETYLNURAMINYLACTOSE-
DE BINDING FIBRILLAR HEMAGGLUTININ RECEPTOR-BINDING SUBUNIT) (NLBH)
DE (FLAGELLAR SHEATH ADHESIN) (ADHESIN A) (FRAGMENT).
DE HPAa OR HNAa.
OS Helicobacter acinonyx.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96001251.
RA Evans D.G., Lampert H.C., Nakano H., Eaton K.A., Burnens A.P.,
RA Brionsdon M.A., Evans D.J. Jr.

```


RT "Genetic evidence for host specificity in the adhesin-encoding genes
 RT hxaA of Helicobacter acinonyx, hnaA of H. nemeurine and hpaA of H.
 RT pylori".
 RL Gene 163:97-102(1995).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
 CC ANCHOR (PROBABLY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U27126; AAA85563.1; -
 CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; PARTIAL.
 CC KW Flagella; Outer membrane; Lipoprotein.
 CC FT NON_TER 92 97 N-ACETYL-NEURAMINYL-ALPHA(2,3)-LACTOSE
 CC FT BINDING MOTIF (POTENTIAL).
 CC FT NON_TER 125 125
 CC FT SEQUENCE 125 AA; 13923 MW; 8060097C2E6573E2 CRC64;
 CC -----
 CC Query Match 48.2%; Score 771; DB 1; Length 125;
 CC Best Local Similarity 85.6%; Pred. No. 5.54e-118;
 CC Matches 107; Conservative 14; Mismatches 4; Indels 0; Gaps 0;
 CC -----
 CC Db 1 NHPASETIOALDENTILLKPAFOYSDNVAKEKNCNOIALKVEIILNOGKRVISD 60
 CC 43 NHPASEKVOALDEKILLRPAFOYSDNIAKEKNCNOIALKVEIILNOGKRVISD 102
 CC 61 SSDKDDLSRQKKEGVLTLSEGIYLRDPKRTTKQKSGFLFTGDKMGVLTISAG 120
 CC 103 SSDKDDLSRQKKEGVLTLSEGIYLRDPKRTTKQKSGFLFTGDKMGVLTISAG 162
 CC 121 FYKVT 125
 CC 163 FIKVT 167
 CC -----
 CC RESULT 7
 CC ID Y329 METAJA STANDARD; PRT; 616 AA.
 CC AC O57775;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
 CC DE HYPOTHEICAL PROTEIN MJ0329.
 CC GN MJ0329
 CC OS Methanococcus jannaschii.
 CC SC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC NC Methanococcus.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 CC RX MEDLINE: 96337999.
 CC RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 CC RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 CC RA Kerevage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 CC RA Overbeek R., Kirkness E.F., Weissstock K.G., Merrick J.M., Glodek A.,
 CC RA Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 CC RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 CC RA Cotton M.D., Roberts C.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 CC RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 CC RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 CC jannaschii.";
 CC RT Science 273:1058-1073(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U67487; AAB98317.1; -
 CC DR TIGR: MJ0329; -
 CC KW Hypothetical protein.
 CC SO SEQUENCE 616 AA; 72036 MW; 2991EC36DAD64B82 CRC64;
 CC -----
 CC Query Match 6.8%; Score 109; DB 1; Length 616;
 CC Best Local Similarity 25.7%; Pred. No. 2.68e-01;
 CC Matches 19; Conservative 27; Mismatches 23; Indels 5; Gaps 5;
 CC -----
 CC Db 363 GELIAAP-DLTKREV-EKLLMAKEK-GLLEYA-DEEOKLLKRSLEKISK-LE 417
 CC 174 GESLDSTMDLSEIDIOEKLTHTSHSGGLVSTWYKGTNSDAIKRALNIFANIMQ 233
 CC 418 KVGDKFGKSMENF 431
 CC 234 EIDKRLQKLTSEY 247
 CC -----
 CC RESULT 8
 CC ID ORCS-YEAST STANDARD; PRT; 616 AA.
 CC AC P54790;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE ORIGIN RECOGNITION COMPLEX SUBUNIT 3 (ORIGIN RECOGNITION COMPLEX
 CC PROTEIN 62 KDA SUBUNIT).
 CC GN ORC3 OR OAR1 OR OIE1 OR YLL004W OR L1365.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 CC OC Saccharomycetaceae; Saccharomyces.
 CC RN [1]
 CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 CC RX MEDLINE: 96069857.
 CC RA Bell S.P., Mitchell J., Leber J., Kobayashi R., Stillman B.;
 CC RT "The multidomain structure of Orcp reveals similarity to regulators
 CC of DNA replication and transcriptional silencing.";
 CC RL Cell 83:563-568(1995).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 96239540.
 CC RA Hardy C.F.;
 CC RT "Characterization of an essential Orc2p-associated factor that plays
 CC a role in DNA replication.";
 CC RL Mol. Cell. Biol. 16:1832-1841(1996).
 CC RN [3]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-S288C / FY23;
 CC RX MEDLINE: 96405918.
 CC RA Miosga T., Zimmermann F.K.;
 CC RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
 CC a 43.7 kb fragment of chromosome XII including an open reading frame
 CC homologous to the human cystic fibrosis transmembrane conductance
 CC regulator protein CTR.";
 CC RL Yeast 12:693-708(1996).
 CC -----
 CC -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
 CC BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
 CC REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
 CC THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION.
 CC -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS OF 120 KDA, 71 KDA, 62
 CC KDA, 56 KDA, 53 KDA AND 50 KDA.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG

```

CC BINDING DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC P3C SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Y12002; CAAT2731.1; -.
CC DR PFAM: PF00130; DAG_PE-bind; 2.
CC DR PFAM: PF00069; Kinase; 1.
CC DR PFAM: PF00433; Kinase; C; 1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE: PS00479; DAG_PE_BINDING_DOMAIN; 2.
CC DR Transferase: Serine/threonine-protein kinase; ATP-binding; Zinc;
CC KM Phorbol-ester binding; Duplication.
CC -----
CC DOMAIN 458 505 PHORBOL-ESTER AND DAG BINDING (BY
CC SIMILARITY).
CC PHORBOL-ESTER AND DAG BINDING (BY
CC SIMILARITY).
CC FT DOMAIN 526 576 PHORBOL-ESTER AND DAG BINDING (BY
CC SIMILARITY).
CC FT DOMAIN 670 679 POLY-ALA.
CC FT DOMAIN 714 717 POLY-GLN.
CC FT DOMAIN 746 754 POLY-GLN.
CC FT DOMAIN 817 1076 PROTEIN KINASE.
CC FT NP_BIND 823 831 ATP (BY SIMILARITY).
CC FT BINDING 846 846 ATP (BY SIMILARITY).
CC FT ACT_SITE 942 942 BY SIMILARITY.
CC SQ SEQUENCE 1142 AA; 127897 MW; C599012DIDIF5970 CRC64;

Query Match 6.4%; Score 102; DB 1; Length 1142;
Best Local Similarity 23.2%; Pred. No. 1.76e+00;
Matches 22; Conservative 28; Mismatches 39; Indels 6; Gaps 6;

Db 199 AAKRVS-KKIVLTKAKRYELHIDVDGDDSDINPALKRPSTGLSTRILAV 257
QY 46 PASKVQDLKRLILRLAF-QYSD-NIAKEYNFKNQTLAK-VEQILQNG-GYKIVSY 101
Db 258 KDVPHAPLGRFARSPETFIYAKEDIVARTRPSR 292
QY 102 DSSKDDPS-FAQKKEGYLAVAMNGEIVLRPPDKR 135

RESULT 12
ID BL3 CHLTR STANDARD; PRT: 221 AA.
AC OR4533;
YR 15-FEB-2000 (Rel. 39, Created)
JT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L3.
GN RPLC OR BL3 OR CT528.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UM-3/CX;
RA MEDLINE: 99000809.
RA Stephens R.S., Olinger L., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao O., Koonin E.V.,
RA Davis R.W.;
RT Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.
RL Science 282:754-759(1998).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
CC PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
CC THE RIBOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE001323; AAC68129.1; -.
CC DR PFAM: PF00297; Ribosomal_L3; 1.
CC DR PROSITE: PS00474; RIBOSOMAL_L3; FALSE_NEG.
CC KM Ribosomal protein; rRNA-binding.
CC SQ SEQUENCE 221 AA; 23514 MW; 065BDC1462353F5B CRC64;

Query Match 6.3%; Score 100; DB 1; Length 221;
Best Local Similarity 36.7%; Pred. No. 2.97e+00;
Matches 18; Conservative 13; Mismatches 15; Indels 3; Gaps 3;

Db 23 VTSVDANVVAQLK-TASSDGNAYVOMGADVQAPK-K-TIERKSKALL 74
QY 98 VTSVSDSDDDPSFAQKKEGYLAVAMNGEIVLRPPDKRTRKRSERGL 146

RESULT 13
ID YN7 YEAST STANDARD; PRT: 767 AA.
AC P53852;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE CYSTEINYL-TRNA SYNTHETASE C29E6.06C (EC 6.1.1.16) (CYSTEINE--
DE TRNA LIGASE) (CYRS).
GN YNL247W OR N0885.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-S288C / FY1679;
CC MEDLINE: 97377997.
CC RA Sen Gupta M., Gueidener U., Beinhauer J.D., Fiedler T.A.,
CC RA Hegemann J.H.;
CC RT "Sequence analysis of the 33 kb long region between ORC5 and SUF1
CC RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
CC RL Yeast 13:849-860(1997).
CC RN [2]
CC RP SEQUENCE OF 44-767 FROM N.A.
CC RA Poehlmann R., Philippson P.;
CC RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
CC CC -1- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) -> AMP +
CC CC PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).
CC CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X96722; CA965497.1; -.
CC DR PFAM: 271523; CA965497.1; -.
CC DR PFAM: PF01406; tRNA-synt_1e; 1.
CC DR PRINTS: PRO0983; TRNASYNTHCYS.
CC DR PROSITE: PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
CC KM Hypothetical protein; Aminoacyl-tRNA synthetase; Protein biosynthesis;
CC KM Ligase; ATP-binding.
CC FT SIMILAR 65 75 "HIGH" REGION.
CC FT SIMILAR 427 431 "KMSK" REGION.
CC FT BINDING 430 430 ATP (BY SIMILARITY).
CC SQ SEQUENCE 767 AA; 87530 MW; 495D526781B4BE2B CRC64;

Query Match 6.3%; Score 101; DB 1; Length 767;

```

Best Local Similarity 27.7%; Pred. No. 2.29e+00;
Matches 18; Conservative 16; Mismatches 26; Indels 5; Gaps 4;

DB 101 QNVTIDDKIIRARONILE-DNEFKENDTKENAVVDKVKALFQYINKNFITGSEIK 159
QY 49 EKVQALDEKILL-LRPAFQYSDNIKEKFNQKALKEQIL-Q--NGYKIVSDSS 104
DB 160 TEEF 164
QY 105 DKDDF 109

RESULT 14
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC Q00799;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
GN RBP2.
OS Plasmodium vivax (strain Belm).
NC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92315338.
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M88098; AAA29744.1; -
KW Malaria; Receptor; Membrane.
-RT NON_TER 1
T 1251 1251
Q SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 6.3%; Score 100; DB 1; Length 1251;
Best Local Similarity 29.9%; Pred. No. 2.97e+00;
Matches 20; Conservative 17; Mismatches 26; Indels 4; Gaps 4;

DB 759 FESKNNVLETFENMSKNTMLDVHKNID-AYKV-ALEIIAHSDEIDTKOKDSKL-IEK 815
QY 65 FOYSDNIKEKFNQKALKEQILQNGYKIVSDS-SDKDDPSFAQKKEGYLAVALM 123
DB 816 GNOIYK 822
QY 124 NGIYLR 130

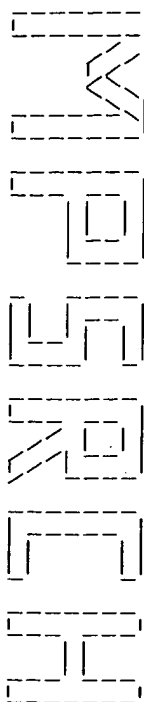
RESULT 15
ID YUL3_YEAST STANDARD; PRT; 1803 AA.
AC P47024; P87192;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON TY4 207.7 KDA HYPOTHETICAL PROTEIN.
GN TY4B OR YUL113M OR J0780.
OS Saccharomyces cerevisiae (Baker's Yeast).
NC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-S288C / FY1679;
RX MEDLINE; 97103775.
RA Czepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, COT7, GZP3, two tRNA genes,
three remnant delta elements and a Ty4 transposon";
RL Yeast 12:1471-1474(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z49389; CAA89409.1; -
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284A8D5D23 CRC64;

Query Match 6.3%; Score 100; DB 1; Length 1803;
Best Local Similarity 24.4%; Pred. No. 2.97e+00;
Matches 20; Conservative 28; Mismatches 30; Indels 4; Gaps 4;

DB 1171 NKEELKVTDKNPPTDNGTNSPRLTEONIEAGSPVOTVNSAFINKESSLNKKRKRH 1230
QY 183 DSESLIQEKFLEKTTSHSGGLVSTMGKGTDSNDAL-KRA-LNKIFANI-HQEIDKTL 239
DB 1231 DKNSLSTSELEDKRKRK 1252
QY 240 TQKN-LESYQKAKELKGRNR 260

Search completed: Tue Sep 26 16:20:13 2000
Job time : 11 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch.PP protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Sep 26 16:15:28 2000; Maspar time 15.30 Seconds
Tabular output not generated. 718.193 Million cell updates/sec

Title: >US-09-308-435-2
Description: (28-260) from US09308435.pap
Perfect Score: 1594
Sequence: 1 CSPHIETNEVAKLNKYNHPA.....QKNLESTQKDAKELKGRNR 233

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r64
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 47.512; Variance 117.669; scale 0.404

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------------|-----------|
| 1 | 1580 | 99.1 | 260 | 2 | E64619 | 1.16e-237 |
| 2 | 1564 | 99.1 | 260 | 2 | neuraminylase-bln | 6.15e-235 |
| 3 | 843 | 52.9 | 124 | 2 | PC4089 | 2.25e-113 |
| 4 | 766 | 48.1 | 124 | 2 | PC4090 | 1.25e-100 |
| 5 | 230 | 14.4 | 249 | 2 | E71865 | 4.34e-16 |
| 6 | 222 | 13.9 | 249 | 2 | B64571 | 6.09e-15 |
| 7 | 171 | 10.7 | 282 | 2 | A71930 | 5.00e-08 |
| 8 | 171 | 10.1 | 278 | 2 | D64581 | 9.69e-07 |
| 9 | 110 | 6.9 | 1802 | 2 | S52611 | 1.00e+00 |
| 10 | 108 | 6.8 | 320 | 2 | F75063 | 1.63e+00 |
| 11 | 107 | 6.7 | 616 | 2 | S64746 | 2.07e+00 |
| 12 | 106 | 6.6 | 394 | 2 | E70135 | 2.63e+00 |
| 13 | 105 | 6.6 | 805 | 2 | S73374 | 3.33e+00 |
| 14 | 105 | 6.6 | 1465 | 2 | S31262 | 3.33e+00 |
| 15 | 105 | 6.6 | 1803 | 2 | S56894 | 3.33e+00 |
| 16 | 104 | 6.5 | 313 | 1 | XYNHCL | 4.22e+00 |
| 17 | 104 | 6.5 | 417 | 2 | T15174 | 4.22e+00 |
| 18 | 103 | 6.5 | 527 | 2 | JC1267 | 5.34e+00 |
| 19 | 104 | 6.5 | 616 | 2 | A63341 | 4.22e+00 |
| 20 | 102 | 6.4 | 559 | 2 | T08523 | 6.74e+00 |
| 21 | 102 | 6.4 | 559 | 2 | S31177 | 6.74e+00 |
| 22 | 100 | 6.3 | 221 | 2 | F71507 | 1.07e+01 |
| 23 | 101 | 6.3 | 319 | 2 | E70091 | 8.50e+00 |

| | | | | | | |
|----|-----|-----|------|---|--------|-----------------------|
| 24 | 101 | 6.3 | 471 | 2 | S54460 | hypothetical protein |
| 25 | 101 | 6.3 | 571 | 2 | F70040 | sulfite reductase hom |
| 26 | 101 | 6.3 | 592 | 2 | G69159 | sensory transduction |
| 27 | 101 | 6.3 | 767 | 2 | S63220 | probable membrane pro |
| 28 | 101 | 6.3 | 856 | 2 | C71956 | probable endopeptidas |
| 29 | 101 | 6.3 | 1000 | 2 | T13636 | probable minor struct |
| 30 | 100 | 6.3 | 1252 | 2 | B42771 | reticulocyte-binding |
| 31 | 99 | 6.2 | 405 | 2 | T10260 | patatin-like protein |
| 32 | 98 | 6.2 | 1134 | 2 | D75014 | hypothetical protein |
| 33 | 99 | 6.2 | 1467 | 2 | PC1253 | tyb protein - yeast (|
| 34 | 98 | 6.1 | 119 | 2 | PL0089 | ig heavy chain V reg1 |
| 35 | 97 | 6.1 | 308 | 2 | C75125 | gmp synthase, cter do |
| 36 | 98 | 6.1 | 311 | 2 | S73625 | methionyl-tRNA formyl |
| 37 | 98 | 6.1 | 330 | 2 | S58255 | hypothetical protein |
| 38 | 98 | 6.1 | 705 | 2 | S54521 | probable membrane pro |
| 39 | 98 | 6.1 | 764 | 2 | B64570 | phenylalanine--trna 1 |
| 40 | 98 | 6.1 | 980 | 2 | S54986 | regulatory protein - |
| 41 | 97 | 6.1 | 9376 | 2 | T14593 | syringomycin syntheta |
| 42 | 96 | 6.0 | 231 | 2 | D64485 | hypothetical protein |
| 43 | 96 | 6.0 | 368 | 2 | A28825 | keratin, type I nonep |
| 44 | 96 | 6.0 | 672 | 2 | F65096 | probable NADH-depende |
| 45 | 96 | 6.0 | 1265 | 2 | F71429 | hypothetical protein |

ALIGNMENTS

| RESULT | 1 | ENTRY | ALTERNATE_NAMES | ORGANISM | DATE | ACCESSIONS | REFERENCE | #authors |
|--------|---|--|-----------------|----------|------|-------------|--------------------------------|--------------|
| | | E64619 | #type complete | | | E64619 | | |
| | | adhesin A - Helicobacter pylori (strains 26695 and 8826) | | | | 09-Aug-1997 | #sequence_revision 09-Aug-1997 | #text_change |
| | | flagellar sheath adhesin hpa1; | | | | 26-Feb-1999 | | |
| | | N-acetylneuraminylactose-binding fibrillar hemagglutinin | | | | | | |
| | | receptor-binding subunit | | | | | | |
| | | Receptor-binding subunit | | | | | | |
| | | TomB, J.F.; White, O.; Karlavage, A.R.; Clayton, R.A.; | | | | E64619 | B47052; PC4088 | |
| | | Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klein, | | | | A64520 | | |
| | | H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, | | | | | | |
| | | J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; | | | | | | |
| | | Richardson, D.; Dodson, R.; Khalak, H.G.; Glodok, A.; | | | | | | |
| | | McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; | | | | | | |
| | | Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Uterback, T.R.; | | | | | | |
| | | Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Waldman, J.M.; | | | | | | |
| | | Fujii, C.; Bowman, C.; Matthey, L.; Wallin, E.; Hayes, | | | | | | |
| | | W.S.; Borodovsky, M.; Karpk, F.D.; Smith, H.O.; Fraser, | | | | | | |
| | | C.M.; Venter, J.C. | | | | | | |
| | | Nature (1997) 388:539-547 | | | | | | |
| | | The complete genome sequence of the gastric pathogen | | | | | | |
| | | Helicobacter pylori. | | | | | | |
| | | #cross-references MIM:97394467 | | | | | | |
| | | E64619 | | | | | | |
| | | #status | | | | | | |
| | | nucleic acid sequence not shown; translation not shown | | | | | | |
| | | #molecule_type DNA | | | | | | |
| | | 1-260 #label TOM | | | | | | |
| | | #cross-references GB:AE000591; GB:AE000511; NID:g2313918; PID:g2313925; | | | | | | |
| | | TIGR:HP0797 | | | | | | |
| | | ##experimental_source strain 26695 | | | | | | |
| | | REFERENCE | | | | | | |
| | | A47052 | | | | | | |
| | | #authors | | | | | | |
| | | Evans, D.G.; Karjalainen, T.K.; Evans Jr., D.J.; Graham, | | | | | | |
| | | D.Y.; Lee, C.H. | | | | | | |
| | | J. Bacteriol. (1993) 175:674-683 | | | | | | |
| | | #journal | | | | | | |
| | | Cloning, nucleotide sequence, and expression of a gene | | | | | | |
| | | encoding an adhesin subunit protein of Helicobacter pylori. | | | | | | |
| | | #cross-references MIM:93139035 | | | | | | |
| | | #accession | | | | | | |
| | | B47052 | | | | | | |
| | | #status | | | | | | |
| | | preliminary | | | | | | |
| | | #molecule_type DNA | | | | | | |
| | | 1-2, 'r', '4', 'G', '6-18', 'T', '20-61', 'K', '63-83', 'T', '85-88', 'E', | | | | | | |
| | | '90-99', 'N', '101-108', 'F', '110-111', 'A', '113-152', 'L', '154-179, | | | | | | |
| | | 'LEMI', #label EVAL | | | | | | |
| | | #cross-references GB:X61574; NID:g732735 | | | | | | |

```

#note extracted from NCBI backbone (NCBI:123532,
#accession C47052
#status preliminary
#molecule-type DNA
#residues 'NM', 204-218, 'R', 220, 'LC', 223-228, 'RS', 231-234, 'M',
#cross-references GB:X61574; MID:9732735
#note sequence extracted from NCBI backbone (NCBI:123532,
NCBI:123535)
PC4088
#authors Evans, D.G.; Lamperl, H.C.; Nakano, H.; Eaton, K.A.; Burnens
A.P.; Bronston, M.A.; Evans Jr., D.J.
#journal Gene (1995) 163:97-102
#title Genetic evidence for host specificity in the adhesin-encoding
genes hxa of Helicobacter acinonyx, hnaa of H. nemestriana
and hpa of H. pylori.
#cross-references MIM:2601251
#accession PC4088
#molecule-type DNA
#residues 43-61, 'K', 63-83, 'T', 85-88, 'E', 90-99, 'N', 101-108, 'F',
#experimental_source strain 8826 110-111, 'A', 113-123, 'I', 125-166 #label EVA3
#feature hpa: HP0797
#gene hpa: HP0797
#feature hpa: HP0797
#length 260 #molecular-weight 29040 #checksum 2565
#region receptor binding #status predicted
#length 260 #molecular-weight 29040 #checksum 2565
Query Match 99.1%; Score 1580; DB 2; Length 260;
Best Local Similarity 98.7%; Pred. No. 1.16e-237;
Matches 230; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 28 CSPHIETNEVALNTNYPASEKQALDEKILLRPAFOYSDIAKEYENKFNKTALKV 87
Oy 28 CSPHIETNEVALNTNYPASEKQALDEKILLRPAFOYSDIAKEYENKFNKTALKV 87
Db 88 EQLIANGYKIVSYDSSDKDLSFSQKEGYLAVANGELVLRPDKRTIQKSEGLF 147
Oy 88 EQLIANGYKIVSYDSSDKDLSFSQKEGYLAVANGELVLRPDKRTIQKSEGLF 147
Oy 88 EQLIANGYKIVSYDSSDKDLSFSQKEGYLAVANGELVLRPDKRTIQKSEGLF 147
Db 148 STGDKKMEGYLIPAGFYKTIILPEMSESLDSTMDLSLIDQEKLTTHSHSGGLYS 207
Oy 148 STGDKKMEGYLIPAGFYKTIILPEMSESLDSTMDLSLIDQEKLTTHSHSGGLYS 207
Oy 148 STGDKKMEGYLIPAGFYKTIILPEMSESLDSTMDLSLIDQEKLTTHSHSGGLYS 207
Db 208 TMVGTGNSNDIAKSLNKIFANIMOBIDKLTQKNLESYQDAKELKGRNR 260
Oy 208 TMVGTGNSNDIAKSLNKIFANIMOBIDKLTQKNLESYQDAKELKGRNR 260
Oy 208 TMVGTGNSNDIAKSLNKIFANIMOBIDKLTQKNLESYQDAKELKGRNR 260
ESQLT 2 C71896 #type complete
ENTRY C71896 #type complete
TITLE neuraminyllactose-binding hemagglutinin precursor -
ORGANISM Helicobacter pylori (strain J99)
#formal_name Helicobacter pylori
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
#variety strain J99
#accession C71896
#authors Alm, R.A.; Ling, L.S.T.; Molt, D.T.; King, B.L.; Brown, E.D.;
Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deGongle,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MIM:99120557
#accession C71896
#status preliminary

```

[illegible]

[illegible]

```
#title      The complete genome sequence of the gastric pathogen  
#            Helicobacter pylori.  
#cross-references XREF:9739467  
#accession  D64581  
#status     preliminary; nucleic acid sequence not shown;  
            translation not shown  
#molecule-type DNA  
##residues  1-278 #label TOM  
##cross-references GB:AE000564; GB:AE000511; NID:g2313602; PID:g2313613  
            TIGR:HP0492  
SUMMARY     #length 278 #molecular-weight 31947 #checksum 4354  
  
Query Match          10.1%; Score 161; DB 2; Length 278;  
Best Local Similarity 25.0%; Pred. No. 9,69e+07;  
Matches 36; Conservative 61; Mismatches 85; Indels 22; Gaps 18;  
  
Db 62 PLDFNF-P-I-HIYAQONHHVGIIPRIQVSDNL-KPYIDKFDALINQITFEKRGY 118  
   :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 39 ALKLNHNPASEKQA-IDEKIL-LLPAPFYSDNIAKEYENKKNQALAKVEQILQNQGY 96  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 119 QVLRFOD-EKA-LMAQDKRKIFSVLDLKGWGLIEDLKNNL--K-DPN---NPMLDTL-- 168  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 97 KYIVSDSKDPEFSFAOKKEGYLAVMNGEIVLRPPDKRTIQKKSEGLLEFSTLDKMEG 156  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 169 VDOSSGVWFNYEPESNRVHDFAVEGTFOAM-TY-TYKH-NNSGGLNSSNIIEHYL 225  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 157 VIIPAFIFVTLLTEPMSGESLDSFTMDLSELDIOEFKLTKTHSSHSGGL-VS-TMWKG-T 213  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 226 ENKEDAIHKILNBMYAVYMKRAVETLTENIDKYREADIRMG 269  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 214 D-NSDAIKSALKNTIPANTMOEIDKTLQKNESYQDAKELKG 256  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
  
RESULT      9  
ENTRY       S52611 #type complete  
TITLE       Tyb protein - yeast (Saccharomyces cerevisiae)  
            retrotransposon Ty4  
ALTERNATE_NAMES  
ORGANISM    protein YHL008w-a  
DATE        #formal_name Saccharomyces cerevisiae  
            05-May-1995 #sequence_revision 19-Oct-1995 #text_change  
            08-Nov-1996  
ACCESSIONS  S52611  
REFERENCE    S46794  
AUTHORS      Favello, T.  
SUBMISSION  submitted to the EMBL Data Library, June 1994  
DESCRIPTION  The sequence of S. cerevisiae cosmid L5018.  
#accession   S52611  
#molecule-type DNA  
##residues   1-1802 #label FAV  
##cross-references EMBL:U11581  
GENETICS  
#map_position BL  
#mobile_element retrotransposon Ty4  
SUMMARY     #length 1802 #molecular-weight 207968 #checksum 3295  
  
Query Match          6.9%; Score 110; DB 2; Length 1802;  
Best Local Similarity 26.8%; Pred. No. 1.00e+00;  
Matches 22; Conservative 27; Mismatches 29; Indels 4; Gaps 4;  
  
Db 1170 NLKELKVTDKNVPYDGNVSPRLQNIASGPSVOTVNRSAFLNKFFSLNKKRRKRH 1229  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 183 DLSELDDIOEKPLKTHSHSGGLVSTMVGTGDNNDAT-KSA-LNKIFANI-MOEIDKTL 239  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 1230 DKNNLSLYSELERDKRSKRNR 1251  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Qy 240 TOKN-LESYQDAKELKGKRN 260  
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
  
RESULT      10  
ENTRY       F75063 #type complete  
TITLE       hypothetical protein PAB0993 - Pyrococcus abyssi (strain  
            Jorsay)
```


DATE 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change

ACCESSIONS F75063

REFERENCE A75001

#authors anonymous, Genoscope

#submission submitted to the EMBL Data Library, July 1999

#description Pyrococcus abyssi genome sequence; insights into archaeal chromosome structure and evolution.

#accession F75063

#status preliminary

#molecule_type DNA

#residues 1-320 ##label RAW

#cross-references GB:AJ248287; GB:AL096836; NID:95458657; PIDN:CA50403.1; PID:e1516301; PID:95458916

#experimental_source strain Orsay

GENETICS

#gene PAB0993

SUMMARY #length 320 #molecular-weight 37677 #checksum 8652

Query Match 6.8%; Score 108; DB 2; Length 320;

Best Local Similarity 24.7%; Pred. No. 1.63e+00;

Matches 21; Conservative 32; Mismatches 25; Indels 7; Gaps 7;

Db 131 ERMSFPLEIAE-SLEKRLIKSLEK-RVEELVEKYNNSNGISPEVIREPIDK-YDSLV 187

175 ESLSSTMDISELDIOKFKTKTHSHSGISGVSTMTVGTGNS-N-DAIKSALNKIFANIM 232

QY 188 RE-NVEL-RRLESREKIKDLREK 210

1 233 QEDIKRLTKGNLESYOKDAKELKGR 257

RESULT 11

ENTRY 564746 #type complete

TITLE ORC3 protein - yeast (Saccharomyces cerevisiae)

ALTERNATE_NAMES protein I1365; protein YLL004w

ORGANISM #formal_name Saccharomyces cerevisiae

DATE 01-Aug-1995 #sequence_revision 24-May-1996 #text_change

ACCESSIONS 564746; S70572

REFERENCE 564743

#authors Miosga, T.; Zimmermann, F.K.

#submission submitted to the Protein Sequence Database, May 1996

#accession 564746

#molecule_type DNA

#residues 1-616 ##label MIO

#cross-references EMBL:Z73109; NID:g1360162; PID:e245444; PID:g1360163; MIPS:YLL004w

REFERENCE 570557

#experimental_source strain S288C

#authors Miosga, T.; Zimmermann, F.K.

#journal Yeast (1996) 12:693-708

#title Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on a 43.7 kb fragment of chromosome XII including an open reading frame homologous to the human cystic fibrosis transmembrane conductance regulator protein CFTR.

#cross-references MIMD:96405918

#accession S70572

#status nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-616 ##label MIM

#cross-references EMBL:X91488; NID:g1495203; PID:e199016; PID:g1495220

#note the nucleotide sequence was submitted to the EMBL Data Library, September 1995

GENETICS

#gene SGD:ORC3; OAF1

#cross-references SGD:S0003927; MIPS:YLL004w

SUMMARY #map_position 12L

#length 616 #molecular-weight 72076 #checksum 76

Query Match 6.7%; Score 107; DB 2; Length 616;

Best Local Similarity 25.7%; Pred. No. 2.07e+00;

Matches 26; Conservative 33; Mismatches 32; Indels 10; Gaps 10;

Db 197 KDLAMVFNFKVDVDSINFNTLDFILKSAFKYDHVKISLFTNNTNLSNIEKNLRQSTI 256

36 NEVALKLNTHPA-SEKQALDEKILLRPAFYSD-NIAKYE-N-KFKN-QTALKVEQI 90

Db 257 RLKRNTHKLDV-SSNKG-FKYNQIFOSFLD-TVDRKML 294

91 -LONOGYKVISVSDSKDDFSFAOKK-EGYLAVAMNGEIVL 129

RESULT 12

ENTRY E70135 #type complete

TITLE flagellar protein (flbc) homolog - Lyme disease spirochete

ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease spirochete

DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change

ACCESSIONS E70135

REFERENCE A70100

#authors Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, T.; Matthey, L.; McDonald, L.; Attach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, R.; Hatch, B.; Smith, H.O.; Venter, J.C. Nature (1997) 390:580-586

#journal Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

#title

#cross-references MIMD:98065943

#accession E70135

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-394 ##label KLE

#cross-references GB:AE001137; GB:AE000783; NID:g2688160; PID:g2688181; TIGR:B80285

#experimental_source strain B31

SUMMARY #length 394 #molecular-weight 45073 #checksum 5653

Query Match 6.8%; Score 106; DB 2; Length 394;

Best Local Similarity 19.7%; Pred. No. 2.63e+00;

Matches 24; Conservative 44; Mismatches 48; Indels 6; Gaps 6;

Db 249 DFVSSLKPEWNLKINKNIYDKA-KVY-LKSNNTGEIKLVLPKPELGSIRINLMLDSNNL 306

QY 74 EYENKFKNOATLKVDEIOLNKGKIVSVSDSDKDDFSFAOK-KE-GYLAVAMNGEIVLRP 131

Db 307 LGRIVVDNONVR-MFDONNHSINKMLGESGF-NASLNTFLAGENTSFTGNFKRDKSDQ 364

QY 132 DPKRTIOKSEPLRSTGDKMEGVLPAGFIKVTILPEMSESDSFMDSELDIOE 191

Db 365 NF 366

QY 192 KF 193

RESULT 13

ENTRY S73374 #type complete

TITLE phenylalanine--tRNA ligase (Pc 6.1.1.20) beta chain phet - Mycoplasma pneumoniae (ATCC 29342) (SGC3)

ALTERNATE_NAMES hypothetical protein C09_orf805; phenylalanyl-tRNA synthetase beta chain phet

ORGANISM #formal_name Mycoplasma pneumoniae

#variety ATCC 29342

DATE 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change

ACCESSIONS S73374

REFERENCE S73327

#authors Himmelreich, R.; Hilbert, H.; Piagens, H.; Pirkl, E.; Li,

```

#journal      B.C.; Herrmann, R.
#title        Nucleic Acids Res. (1996) 24:4420-4449
#cross-ref    Complete sequence analysis of the genome of the bacterium
#accession    Mycoplasma pneumoniae.
#cross-ref    M01D:97105885
#accession    S73374
#status       preliminary; nucleic acid sequence not shown;
#molecule    translation not shown
#residues     1-805 ##label HIM
#cross-ref    EMBL:AE000006; GB:U00089; NID:91673695;
#note         PIDN:AB95696.1; PID:91673698
#note         the nucleotide sequence was submitted to the EMBL Data
#note         Library, November 1996

GENETICS
#gene          phet
#genetic_code  SGC3
#classification #superfamily phenylalanine--tRNA ligase beta chain
#keywords      aminocyl-tRNA synthetase; ligase; protein biosynthesis
SUMMARY
#length 805 #molecular-weight 91713 #checksum 4320

Query Match
Best Local Similarity 28.4%; Score 105; DB 2; Length 805;
Matches 23; Conservative 21; Mismatches 33; Indels 4; Gaps 4;

Db 506 ALQKRTKIQTGPHNVITYQLISPERANFELGSLNWEIKNP-LSNERSYLKGLID 564
QY 150 GIDKMEGVLPAGFIVTILEPMSGESLDSFTW-DISEL-DIQEKFLKTHSHSGGLVS 207
Db 565 SLRVIQK-NAAYKKNLGNIF 584
QY 208 TWKGTGTDNSDAIKSALNKIF 228

RESULT 14
ENTRY      S31262 #type fragment
TITLE      Tyb protein - yeast (Saccharomyces cerevisiae)
ORGANISM   retrotransposon Ty4 (fragment)
#formal_name Saccharomyces cerevisiae
DATE       18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change
25-Apr-1997
ACCESSIONS S31262
REFERENCE   S31261
#authors    Janetzký, B.; Lehle, L.
#journal    J. Biol. Chem. (1992) 267:19798-19805
#title      Ty4, a new retrotransposon from Saccharomyces cerevisiae,
#title      flanked by tau-elements.
#cross-ref  M01D:93015829
#accession  S31262
#molecule  type DNA
#residues   1-1465 ##label JAN
#cross-ref  EMBL:S46865
GENETICS
#mobile_element retrotransposon Ty4
SUMMARY
#length 1465 #checksum 1908

Query Match
Best Local Similarity 25.6%; Score 105; DB 2; Length 1465;
Matches 21; Conservative 28; Mismatches 29; Indels 4; Gaps 4;

Db 833 NIKELKVTDKNVPDNGTIVSPRLBNIEAGSPVQTVKSAFLNKEFSSLNKKRRKRH 892
QY 183 DLSELDIOEKFLKTHSHSGGLVSTWVGTDNSDAI-KSA-LNKIFANI-MQETDKL 239
Db 893 DKNSLTSELERDKRKRKR 914
QY 240 TOKN-LESYQDAKELKGRNR 260

RESULT 15
ENTRY      S56894 #type complete
TITLE      Tyb protein - yeast (Saccharomyces cerevisiae)
#mobile_element retrotransposon Ty4.JL

```

```

ALTERNATE_NAMES protein J0780; protein YJL113w
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
14-Nov-1997
ACCESSIONS        S56891
REFERENCE         S56894
#authors          Celeplich, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
#submission       submitted to the Protein Sequence Database, September 1995
#accession        S56894
#molecule        type DNA
#residues         1-1803 ##label CZI
#cross-ref         EMBL:Z49389; MIPS:YJL113w
GENETICS
#map_position 10L
#mobile_element retrotransposon Ty4.JL
SUMMARY
#length 1803 #molecular-weight 207708 #checksum 3450

Query Match
Best Local Similarity 25.6%; Score 105; DB 2; Length 1803;
Matches 21; Conservative 28; Mismatches 29; Indels 4; Gaps 4;

Db 1171 NIKELKVTDKNVPDNGTIVSPRLBNIEAGSPVQTVKSAFLNKEFSSLNKKRRKRH 1230
QY 183 DLSELDIOEKFLKTHSHSGGLVSTWVGTDNSDAI-KSA-LNKIFANI-MQETDKL 239
Db 1231 DKNSLTSELERDKRKRKR 1252
QY 240 TOKN-LESYQDAKELKGRNR 260

```

Search completed: Tue Sep 26 16:15:47 2000
Job time : 19 secs.

 W I T E N E S S

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:14:59 2000; Maspar time 9.09 Seconds

Tabular output not generated. 606.971 Million cell updates/sec

Title: >US-09-308-435-2
 Description: (28-260) from US09308435.pep
 Perfect Score: 1594
 Sequence: 1 CSPHIETNEVALKLNYPHA.....QKNLESYQKDAKELKGRNR 233

Scoring table:
 PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseq36
 1:geneseqp

Statistics: Mean 32.420; Variance 175.543; scale 0.185

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | Pred. No. |
|------------|-------|-------------|--------|-------|-------------------------------|-----------|
| 1 | 1594 | 100.0 | 260 | 1 | W60156 Helicobacter pylori 29 | 4.75e-113 |
| 2 | 1594 | 100.0 | 260 | 1 | W07446 Helicobacter pylori su | 4.75e-113 |
| 3 | 1594 | 100.0 | 461 | 1 | W74466 Adhesin/V.cholerae tox | 4.75e-113 |
| 4 | 1589 | 99.7 | 260 | 1 | W60157 Helicobacter pylori 29 | 1.18e-112 |
| 5 | 1589 | 99.7 | 260 | 1 | W07450 Helicobacter pylori su | 1.18e-112 |
| 6 | 1552 | 97.4 | 260 | 1 | Y11000 H. pylori ORF 11ap2071 | 9.72e-110 |
| 7 | 1545 | 96.9 | 268 | 1 | W20973 H. pylori derived prot | 3.46e-109 |
| 8 | 302 | 18.9 | 97 | 1 | W24663 H. pylori derived prot | 5.42e-13 |
| 9 | 302 | 18.9 | 97 | 1 | W20454 H. pylori ORF 01a61101 | 8.55e-08 |
| 10 | 230 | 14.4 | 249 | 1 | W55445 H. pylori ORF 01a61101 | 8.55e-08 |
| 11 | 230 | 14.4 | 261 | 1 | W55332 H. pylori ORF 04e61110 | 3.15e-07 |
| 12 | 222 | 13.9 | 249 | 1 | W98333 H. pylori ORF 04e61110 | 1.10e-03 |
| 13 | 171 | 10.7 | 282 | 1 | Y10951 H. pylori secreted or | 3.33e+00 |
| 14 | 171 | 10.7 | 285 | 1 | W20938 H. pylori secreted or | 1.21e+01 |
| 15 | 118 | 7.4 | 147 | 1 | W21638 Grapevine leafroll vir | 1.60e+01 |
| 16 | 109 | 6.7 | 615 | 1 | R77276 ORC3 subunit of yeast | 1.60e+01 |
| 17 | 107 | 6.7 | 615 | 1 | W23226 S. cerevisiae orixin o | 1.60e+01 |
| 18 | 107 | 6.7 | 615 | 1 | W23226 Human heart muscle spe | 3.23e+01 |
| 19 | 102 | 6.4 | 552 | 1 | W90172 Helicobacter c1pb (N-t | 4.25e+01 |
| 20 | 100 | 6.3 | 751 | 1 | W13491 Merocaine apical-end p | 4.25e+01 |
| 21 | 100 | 6.3 | 1254 | 1 | R07503 Merocaine apical-end p | 4.25e+01 |
| 22 | 100 | 6.3 | 1254 | 1 | R07503 Pasteurella multocida | 4.88e+01 |
| 23 | 99 | 6.2 | 1285 | 1 | P93089 | |

ALIGNMENTS

| RESULT ID | Score | Query Match | Length | DB ID | Description | Pred. No. |
|-----------|-------|-------------|--------|-------|--------------------------------|-----------|
| 24 | 97 | 6.1 | 119 | 1 | W49813 Amino acid sequence of | 6.42e+01 |
| 25 | 97 | 6.1 | 222 | 1 | R39267 Humanised C4g1 Ig heav | 6.42e+01 |
| 26 | 97 | 6.1 | 222 | 1 | W49817 Fragment of humanised | 6.42e+01 |
| 27 | 97 | 6.1 | 235 | 1 | R39268 Humanised C4g1 Ig heav | 6.42e+01 |
| 28 | 97 | 6.1 | 225 | 1 | W49818 Amino acid sequence of | 6.42e+01 |
| 29 | 97 | 6.1 | 449 | 1 | R43339 Completely humanised C | 6.42e+01 |
| 30 | 97 | 6.1 | 449 | 1 | W49816 Amino acid sequence of | 6.42e+01 |
| 31 | 98 | 6.1 | 764 | 1 | W98549 H. pylori GHPO 1604 pr | 6.42e+01 |
| 32 | 97 | 6.1 | 914 | 1 | W24800 Spinocebellar ataxia | 6.42e+01 |
| 33 | 97 | 6.1 | 1312 | 1 | W33807 Human ataxin-2. | 6.42e+01 |
| 34 | 97 | 6.1 | 1313 | 1 | W60213 Spinocebellar ataxia | 6.42e+01 |
| 35 | 96 | 6.0 | 246 | 1 | R72386 Penunia MADS box trans | 7.36e+01 |
| 36 | 95 | 6.0 | 1140 | 1 | R72386 XAP-1, part of the DNA | 8.43e+01 |
| 37 | 96 | 6.0 | 2039 | 1 | W56322 Haemophilus parvagallin | 7.36e+01 |
| 38 | 96 | 6.0 | 2042 | 1 | W56319 Haemophilus parvagallin | 7.36e+01 |
| 39 | 94 | 5.9 | 119 | 1 | W49814 Amino acid sequence of | 9.65e+01 |
| 40 | 94 | 5.9 | 138 | 1 | R39266 Mouse C4g1 Ig heavy-ch | 9.65e+01 |
| 41 | 94 | 5.9 | 476 | 1 | W02616 Wheat adenylsuccinate | 9.65e+01 |
| 42 | 94 | 5.9 | 1084 | 1 | R71015 Human neuronal calcium | 9.65e+01 |
| 43 | 94 | 5.9 | 1091 | 1 | R33553 Sequence of the alpha | 9.65e+01 |
| 44 | 94 | 5.9 | 1091 | 1 | R33553 Human calcium channel | 9.65e+01 |
| 45 | 94 | 5.9 | 1091 | 1 | R71011 Human neuronal calcium | 9.65e+01 |

RESULT 1
 ID W60156; standard; Protein; 260 AA.

DE 28-SEP-1998 (first entry)

DE Helicobacter pylori 29 kDa surface-exposed antigen.

KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.

PN Helicobacter pylori.

PD WO9822135-A1.

PE 18-MAY-1998.

PF 18-NOV-1997; SE1927.

PR 22-NOV-1996; SE-004296.

PA (ASMR.) ASMPA AB.

PI Berglindh T, Lofroth J;

DR N-PSDB; V34642.

PT Aggregate of negatively charged lipid and Helicobacter pylori

PT antigen - useful in therapeutic or preventative vaccines against

PT both forms of bacterium, to generate systemic and mucosal antibody

PT responses

PS Claim 6; Page 28-30; 51pp; English.

CC This polypeptide comprises a 29 kDa surface antigen which is

CC expressed on the surface of both dividing (bacterial) and resting

CC (colloid) forms of Helicobacter pylori. The antigen gives rise to

CC both systemic and local (mucosal) production of antibodies. It is

CC an adhesin that is conserved in all tested strains of H. pylori,

CC and is also a putative virulence factor. The invention relates to

CC a new pharmaceutical composition comprising a lipid aggregate of a

CC antigenic, native or recombinant antigen of H. pylori, preferably

CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34642)

CC coding for such an antigen. The new formulations, which are useful

CC as vaccine formulations, elicit a protective immune response

CC against H. pylori infections, and are suitable for therapeutic and

CC prophylactic use. An adequate response is produced even without

CC adjuvant. A second 29 kDa antigen (see W60157) is provided that

CC differs only at residue 222 (Arg for Ser).

CC Sequence 260 AA;

SO

Query Match 100.0%; Score 1594; DB 1; Length 260;

Best Local Similarity 100.0%; Pred. No. 4.75e-113;

Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 CSPHIETNEVALKLNYPHASEKQVADDEKILLPAFOYSDNIAKEYNKFNKNOTALKV 87

Oy 28 CSPHIETNEVALKLNYPHASEKQVADDEKILLPAFOYSDNIAKEYNKFNKNOTALKV 87

Db 88 EQLLNGGKIVSDSSKDDFSFAQKKEGYLAVMGEIYLRPPDKRTIOKSEPLLF 147

| QY | 88 | EQLIIONOGYKIVSYSDSDKDEFSFAQKKEGYLAIVAMNGEYTLRPDPKRTIOKKSPEGLF | 147 |
|-----------------------|---|---|-----|
| Db | 148 | SYGLDMKEGYLIPAGIKRTITLPEPMGSELSDFYMDLSLDIQEKFLLKTTSSHGGLVS | 207 |
| QY | 148 | SYGLDMKEGYLIPAGIKRTITLPEPMGSELSDFYMDLSLDIQEKFLLKTTSSHGGLVS | 207 |
| Db | 208 | TMVKGTDNSDAIKSALNKIFANIMOEDKTLQKNLSYQDAKELKGRNR | 260 |
| QY | 208 | TMVKGTDNSDAIKSALNKIFANIMOEDKTLQKNLSYQDAKELKGRNR | 260 |
| RESULT | 2 | | |
| ID | M07449 | standard; Protein; 260 AA. | |
| AC | M07449: | | |
| DE | 04-MAR-1997 | (first entry) | |
| KW | Helicobacter pylori surface-exposed 29 kDa antigen. | | |
| KW | Antigen; adhesion; immunogen; vaccine; immunisation; ulcer; | | |
| KW | therapy; diagnosis. | | |
| OS | Helicobacter pylori strain CCUG 17874 (NTCC 11637). | | |
| FM | Key | Location/Qualifiers | |
| FT | peptide | 1..27 | |
| FT | cleavage_site | /label-Sig_peptide | |
| FT | | 25..28 | |
| FT | | /note="consensus recognition site for signal | |
| FT | protein | peptidase II in prolipoproteins" | |
| FT | | 28..260 | |
| FT | | /label-Mat_protein | |
| FT | | /note="either the uncleaved or mature protein can | |
| FT | | be used in applns. of the invention" | |
| PN | M09638475-A1. | | |
| PD | 05-DEC-1996 | | |
| PR | 03-JUN-1996; | SE0727. | |
| PR | 01-JUN-1995; | SE-002007. | |
| PR | 21-MAR-1996; | SE-001085. | |
| PA | (ASTR) ASTRA AB. | | |
| PI | Bolin I, Svennerholm A; | | |
| DR | WPI: 97-034307/03. | | |
| DR | N-PSDB: T43834. | | |
| PT | Helicobacter pylori surface exposed antigen - useful for treatment, | | |
| PT | prophylaxis or diagnosis of mammalian infections | | |
| PS | Claim 2; Page 34-35; 49pp; English. | | |
| CC | A 29 kDa antigen (M07449) of Helicobacter pylori is an adhesin | | |
| CC | expressed on the surface of resting and dividing forms of the | | |
| CC | bacterium. It is conserved in all tested strains of H. pylori, | | |
| CC | and gives rise to both systemic and local (mucosal) protn. of | | |
| CC | antibodies. Its amino acid sequence was deduced from a cDNA clone | | |
| CC | (T43834); a sequence (M07550) deduced from a second cDNA clone | | |
| CC | (T43835) differs only at position 222 (Arg for Ser). Recombinant | | |
| CC | antigen can be produced in transformed host cells. It is useful | | |
| CC | for the diagnosis and treatment of H. pylori infections and for | | |
| CC | use as a vaccine. | | |
| JQ | Sequence | 260 AA; | |
| Query Match | 100.0%; | Score 1594; DB 1; Length 260; | |
| Best Local Similarity | 100.0%; | Pred. No. 4.75e-113; | |
| Matches | 233; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | |
| Db | 28 | CSPHIEETNEFAALKNYPASEKYOALDEKILLRPAFOYSDNIAKRYENKFNQOTLAKY | 87 |
| QY | 28 | CSPHIEETNEFAALKNYPASEKYOALDEKILLRPAFOYSDNIAKRYENKFNQOTLAKY | 87 |
| Db | 88 | EQLIIONOGYKIVSYSDSDKDEFSFAQKKEGYLAIVAMNGEYTLRPDPKRTIOKKSPEGLF | 147 |
| QY | 88 | EQLIIONOGYKIVSYSDSDKDEFSFAQKKEGYLAIVAMNGEYTLRPDPKRTIOKKSPEGLF | 147 |
| Db | 148 | STGLDMKEGYLIPAGIKRTITLPEPMGSELSDFYMDLSLDIQEKFLLKTTSSHGGLVS | 207 |
| QY | 148 | STGLDMKEGYLIPAGIKRTITLPEPMGSELSDFYMDLSLDIQEKFLLKTTSSHGGLVS | 207 |
| Db | 208 | TMVKGTDNSDAIKSALNKIFANIMOEDKTLQKNLSYQDAKELKGRNR | 260 |
| QY | 208 | TMVKGTDNSDAIKSALNKIFANIMOEDKTLQKNLSYQDAKELKGRNR | 260 |

3
ID W74466 standard; Protein; 461 AA.
AC W74466;
DE 18-MAY-1999 (first entry)
DT Adhesin/V.cholerae toxin A2 and B subunit fusion protein.
KW Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW duodenal ulcer.
OS Chimeric - Helicobacter pylori.
OS Chimeric - Vibrio cholerae.
PN WO9853082-A1.
PD 26-NOV-1998.
PF 21-MAY-1997; 000091.
PR 21-MAY-1997; WO-KR0091.
PA (DAEW-) DAEWOONG PHARM CO LTD.
PI Kim BO, Lee BK, Park SK, Yoon SW, Yu YH;
DR WPI; 98-434925/37.
DR N-PESDB; X21955.
PT Fusion gene consisting of the Helicobacter pylori adhesin gene
PT ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
PT useful in the diagnosis and treatment of H. pylori
PS Claim 2; Page 20-21; 32pp; English.
CC This sequence represents the fusion protein of the invention, between the
CC H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC This protein is used in the diagnosis of and as a vaccine against
CC H. pylori, which is the main cause of gastritis and gastric/duodenal
CC ulcers. It may also be used in the production of an anti-H. pylori
CC antibody. Vectors containing the DNA sequence and E. coli strains
CC containing it form the basis of a process to produce the protein, which
CC consists of culturing the microorganism and obtaining the protein. The
CC protein is an effective vaccine against H. pylori due to its excellent
CC immunogenicity for H. pylori, stability within the stomach environment,
CC and its ability to penetrate through the stomach mucous membrane to
CC stimulate sIgA production.
CC Note: This sequence was indexed from WO9853082, which is the first
CC major country equivalent to KR97059278.
SO Sequence 461 AA;

| Query Match | 100.0% | Score 1594 | DB 1 | Length 461 |
|------------------------------|--------------|---|--------------|------------|
| Best Local Similarity 100.0% | | Ident. No. 4,75e-113 | | |
| Matches 233 | Conservative | 0 | Mismatches 0 | Indels 0 |
| | | | Gaps 0 | |
| Db | 39 | CSPHIETNEVALKLNTPASEKYOALDEKILLRPAFYQSDNIACEYEKFKFNQALKY | 98 | |
| Qy | 28 | CSPHIETNEVALKLNTPASEKYOALDEKILLRPAFYQSDNIACEYEKFKFNQALKY | 87 | |
| Db | 99 | EQLIANGQKRYISVDSSDKDDFSQAQKKEGYLVANNGEIVLPDPKRTIQKSEPELFL | 158 | |
| Qy | 88 | EQLIANGQKRYISVDSSDKDDFSQAQKKEGYLVANNGEIVLPDPKRTIQKSEPELFL | 147 | |
| Db | 159 | STGLDKRKEGYLIPAGFKVTILPEMSEESDFTMDISEIDIEKFKTTHSHSGGLV | 218 | |
| Qy | 148 | STGLDKRKEGYLIPAGFKVTILPEMSEESDFTMDISEIDIEKFKTTHSHSGGLV | 207 | |
| Db | 219 | TWAKGTDSNDIAIKSALNKIFANIMOIEIDKLLTQKNLSEYQDKAKELGKRRN | 271 | |
| Qy | 208 | TWAKGTDSNDIAIKSALNKIFANIMOIEIDKLLTQKNLSEYQDKAKELGKRRN | 260 | |

RESULT 4

ID W60157 standard; Protein; 260 AA.

AC W60157;

DT 28-SEP-1998 (first entry)

DE Helicobacter pylori 29 kDa surface-exposed antigen.

KM Cell surface antigen; adhesin; infection; vaccine; drug delivery.

OS Helicobacter pylori.

PN W09882135-A1.

PD 28-MAY-1998.

PF 18-NOV-1997; SE1927.

PR 22-NOV-1996; SE-004296.

RA (ASTR) ASTRA AB.

PI Berglundh T, Lofroth J;
DR WPI: 98-312174/27.
DR N-PSDB: V34643.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Disclosures: Page 35-36; 51pp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (baccillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34643)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second, claimed 29 kDa antigen (see W60156) is
CC provided that differs only at residue 222 (Ser for Arg).
SQ Sequence 260 AA;
Query Match 99.7%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.18e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 28 CSPHIETNEVALKLNYPHASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 87
QY 28 CSPHIETNEVALKLNYPHASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 87
Db 88 EQLLNOGKYKIVSDSSDKDDSFQAQKGGYLAVANNGEIVLRPPDKRTIOKSEGLLF 147
QY 88 EQLLNOGKYKIVSDSSDKDDSFQAQKGGYLAVANNGEIVLRPPDKRTIOKSEGLLF 147
Db 148 STGLDKMEGVLLPAGFIKTYTLEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
QY 148 STGLDKMEGVLLPAGFIKTYTLEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
Db 208 TMVKGTDNSNDIAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
QY 208 TMVKGTDNSNDIAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
Y 208 TMVKGTDNSNDIAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
RESULT 5
ID W07450 standard; Protein; 260 AA.
AC W07450:
DE 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
OS therapy; diagnosis.
KM Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH Key Location/Qualifiers
FT peptide
FT 1..27
FT /label- Sig_peptide
FT cleavage_site 25..28
FT /note- "consensus recognition site for signal
FT peptidase II in prolipoproteins"
FT 28..260
FT /label- Mat_protein
FT /note- "either the uncleaved or mature protein can
FT be used in appls. of the invention"
PN W09638475-A1.
PD 05-DEC-1996.
PE 03-JUN-1996; SE0727.
PR 01-JUN-1995; SE-002007.
PR 21-MAR-1996; SE-001085.
PA (ASTR) ASTRA AB.
PI Bojin I, Svennerholm A;
DR WPI: 97-034307/03.

DR N-PSDB: T43835.
PT Helicobacter pylori surface exposed antigen - useful for treatment,
PT prophylaxis or diagnosis of mammalian infections
PS Claim 2: Page 37-38; 49pp; English.
CC A 29 kDa antigen (W07450) of Helicobacter pylori is an adhesin
CC expressed on the surface of resting and dividing forms of the
CC bacterium. It is conserved in all tested strains of H. pylori,
CC and gives rise to both systemic and local (mucosal) prodn. of
CC antibodies. Its amino acid sequence was deduced from a cDNA clone
CC (T43835); a sequence (W07549) deduced from a second cDNA clone
CC (T43834) differs only at position 222 (Ser for Arg). Recombinant
CC antigen can be produced in transformed host cells. It is useful
CC for the diagnosis and treatment of H. pylori infections and for
CC use as a vaccine.
SQ Sequence 260 AA;
Query Match 99.7%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.18e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 28 CSPHIETNEVALKLNYPHASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 87
QY 28 CSPHIETNEVALKLNYPHASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 87
Db 88 EQLLNOGKYKIVSDSSDKDDSFQAQKGGYLAVANNGEIVLRPPDKRTIOKSEGLLF 147
QY 88 EQLLNOGKYKIVSDSSDKDDSFQAQKGGYLAVANNGEIVLRPPDKRTIOKSEGLLF 147
Db 148 STGLDKMEGVLLPAGFIKTYTLEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
QY 148 STGLDKMEGVLLPAGFIKTYTLEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
Db 208 TMVKGTDNSNDIAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
QY 208 TMVKGTDNSNDIAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
Y 208 TMVKGTDNSNDIAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
RESULT 6
ID Y11000 standard; Protein; 260 AA.
AC Y11000:
DE 08-JUN-1999 (first entry)
DE H. pylori ORF 11ap20714_4797137_f3_45 cell envelope protein.
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
KM secreted protein; cellular protein.
OS Helicobacter pylori.
PN W09618323-A1.
PD 07-MAY-1998.
PE 28-OCT-1997; U19575.
PR 14-JUL-1997; US-891928.
PR 28-OCT-1996; US-739150.
PR 06-DEC-1996; US-759739.
PI (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI: 98-271811/24.
DR N-PSDB: X30467.
PT Helicobacter pylori nucleic acids and proteins - used to develop
PT products for the detection, prevention and treatment of H. pylori
PT infections
PS Claims 27, 31; Page 207-208; 279pp; English.
CC Recombinant or substantially pure preparations of H. pylori polypeptides
CC are disclosed, together with the nucleic acids encoding them. In all,
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
CC secreted proteins or other cellular proteins. Vaccines containing the
CC nucleic acids or proteins are claimed, as are probes containing at least
CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC for treating or reducing the risk of H. pylori infections, and the
CC probes can be used diagnostically for detecting the presence of
CC Helicobacter in a sample. The products are also of use in screening
CC for compounds having the ability to interfere with the H. pylori life
CC cycle or to inhibit H. pylori infection.
SQ Sequence 260 AA;
Query Match 97.4%; Score 1552; DB 1; Length 260;

QY 148 STGLDKMEGVLLPAGFIKVTIILEPMGSESLDSFTWDLSELDIOEKELKTHSSHSGGLVS 207

208 TPAKGTDSNDIAIKSALNRIIFANIMOETDKLTKOKNIESYQDKAKEIKGKRRR 26CQY

| RESULT | 8 |
|--------|----------------------------------|
| ID | W24663 standard; Protein; 97 AA. |
| AC | W24663; |

DE *H. pylori* derived protein.
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease
KW activation; inhibition; bactericidal

| EM | Key | Location/Qualifiers |
|----|-----------------------------------|---------------------|
| KW | detection; antisense; inhibition. | |
| OS | <i>Helicobacter pylori</i> . | |
| FH | Key | |
| EM | also difference 04 | |

| | | |
|----|---------------|-----------------|
| FI | /note- | "encoded by GAK |
| PN | WO9719098-A1. | |
| PD | 29-MAY-1997. | |
| PF | 15-NOV-1996. | TI18542 |

PR 17-NOV-1995; US-561469
PA (ASTR) ASTRA AB.
PI Smith DH;
WPI: 97-298052/27.

DR N-PsDB; T77481.
PT Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
SS Claim 18; Page 178; 235pb; English.

CC This sequence represents an *H. pylori* derived protein, no further
CC details are given in the specification.
CC Helicobacter pylori has been strongly linked to chronic gastritis;
CC duodenal ulcer disease. The nucleic acid sequences of the inventors

are used to evaluate compounds, especially activators or inhibitors of bacterial life cycle, for the ability to bind an *H. pylori* nucleic acid sequence. The nucleic acid sequences, and corresponding proteins, are also useful for generating vaccines for immunising subjects against *H. pylori*.

pyrrolidone use in detecting the presence of helicobacter species in a sample. Antisense nucleic acid sequences of these sequences are used to inhibit expression of a gene from *Helicobacter* species. H. pylori whole genomic DNA was isolated and nebulized to a median size of 2000 bp. Purified RNA fragments were blunt-ended and ligated into

CC 2000 bp. The linker was ligated into the blunt-ended and ligated to unique
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC complementary to the BstXI-cut PMV vectors, while the overhang is not
CC self-complementary. Therefore the linkers will not concatenate nor
CC will the cut vector re-ligate itself again. The linker-adapters insert
CC

Note: The ORF/protein reference number for this sequence was obtained by searching the NCBI database. The amino acid sequence was derived from the ORF. The ORF was used to construct a series of 20 pMPY vectors to construct a series of 20 pMPY vectors. The purified DNA samples were then sequenced.

```
CC from the related specification, W096640893.
SQ Sequence 97 AA;
Query Match 18.9%; Score 302; DB 1; Length 97;
```

| | | |
|-----------------------|--------|---|
| Best Local Similarity | 97.7%; | Pred. No. 5.42e-13; |
| Matches | 43; | Conservative |
| | 0; | Mismatches |
| | | 1; |
| | | Indels |
| | | 0; |
| | | Gaps |
| | | 0; |
| Db | 53 | CSPHIIETNEVALKINTHPASEKVOALDEKILLRPAFYQSKNI |
| | | 96 |

QY 28 CSPHIETNEVALKINYPHASEKVALDEKILLIRPAFOISDNI 71

| | | |
|-----|---------------------------------|--|
| DEF | H. pylori derived protein | |
| DT | 29-JUL-1997 (first entry) | |
| AC | W20454; | |
| ID | W20454 standard; protein; 97 AA | |

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium, life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 SS *Helicobacter pylori*.

| ID | RESULT | 12 |
|----|---|---|
| AC | W98332 | standard; Protein; 249 AA. |
| AC | W98332; | |
| DT | 31-MAR-1999 | (first entry) |
| DE | H. pylori | GHPD 1615 protein. |
| KW | GHPD protein; Helicobacter | infection; gastroduodenal disease; gastritis, |
| CC | peptic ulcer disease. | |
| OS | Helicobacter | pylori. |
| PN | W09843478-A1. | |
| PD | 08-OCT-1998. | |
| PF | 01-APR-1998; | 006371. |
| PR | 29-JUL-1997; | US-902615. |
| PR | 01-APR-1997; | US-833457. |
| QR | 24-JUN-1997; | US-881227. |
| 1 | (HDM-A) MERIEUX ORAVAX | PASTER MERIEUX SERUMS. |
| PI | (IMMR) MERIEUX ORAVAX | PASTER MERIEUX SERUMS. |
| PI | Al-Garavati A, Kleantchous H, Miller C, Omen RP, Tomb J; | |
| DR | WPI; 98-542293/46. | |
| DR | NEW SDBB; X14051. | |
| PT | new isolated Helicobacter | polynucleotides - used to develop products |
| PT | for the diagnosis, prevention and treatment of Helicobacter | |
| PT | infections and gastrointestinal diseases | |
| PS | Claim 8; Page 455-456; | 2054dp; English. |
| CC | This sequence represents a Helicobacter | pylori GHPD protein of the |
| CC | invention. The polypeptides can be used for preventing or treating | |
| CC | Helicobacter infections, and gastroduodenal diseases associated with | |
| CC | these infections, including acute, chronic, and atrophic gastritis, and | |
| CC | peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be | |
| CC | used for the production of antibodies. The products can also be used for | |
| CC | detection and diagnosis. | |
| SA | Sequence | 249 AA; |

| | | | | |
|-----------------------|--------|---------------------|------------|----------------|
| Query Match | 13.98; | Score 222; | DB 1; | Length 249; |
| Best Local Similarity | 25.08; | Pred. No. 3.15e-07; | | |
| Matches | 58; | Conservative | 74; | Mismatches 77; |
| | | | Indels 23; | Gaps 22 |

| | |
|--------|--|
| RESULT | 13 |
| ID | Y10951 standard; Protein; 282 AA. |
| AC | Y10951; |
| DT | 08-JUN-1999 (first entry) |
| DE | H. pylori ORF 04ee1108_3906963.fl.7 cell envelope protein. |
| RV | Vaccine; probe; diagnostic; ORF; cell envelope protein; |
| KW | secreted protein; cellular protein. |
| OS | Helicobacter pylori. |
| PN | W09818323-A1. |
| PD | 07-MAY-1998. |
| PE | 28-OCT-1997; U19575. |
| PR | 14-JUL-1997; US-891928. |
| PR | 28-OCT-1996; US-739150. |
| PR | 06-DEC-1996; US-755739. |
| PA | (ASTR) ASTRA AB. |
| PI | Alm RA, Smith D; |
| DR | WPI; 98-271811/24. |
| DR | N-PSDB: X30418. |
| PT | Helicobacter pylori nucleic acids and proteins - used to develop |
| PT | products for the detection, prevention and treatment of H. pylori |
| PT | infections |
| PS | Claims 27, 31; Page 151; 279pp; English. |
| CC | Recombinant or substantially pure preparations of H. pylori polypeptides |
| CC | are disclosed, together with the nucleic acids encoding them. In all, |
| CC | 73 ORFs are shown. The proteins are variously cell envelope proteins, |
| CC | secreted proteins or other cellular proteins. Vaccines containing the |
| CC | nucleic acids or proteins are claimed, as are probes containing at least |
| CC | 8 nucleotides from the nucleic acid sequences. The vaccines are useful |
| CC | for treating or reducing the risk of H. pylori infections, and the |
| CC | probes can be used diagnostically for detecting the presence of |
| CC | Helicobacter in a sample. The products are also of use in screening |
| CC | for compounds having the ability to interfere with the H. pylori life |
| CC | cycle or to inhibit H. pylori infection. |
| CC | Sequence 282 AA; |

| | | | | |
|-----------------------|--------|---------------------|------------|----------------|
| Query Match | 10.7%; | Score 171; | DB 1; | Length 282; |
| Best Local Similarity | 25.1%; | Pred. No. 1.10e+03; | | |
| Matches | 56; | Conservative | 58; | Mismatches 91; |
| | | | Indels 18; | Gaps 14 |

| | | |
|----|-----|--|
| Db | 64 | PLDENE-PV-HIYAPONHHVAGLIMPRIOYSDUL-SPYIDKFDODALINDIQTFFEEKRGY 12 |
| QY | 39 | AKKNTYHPASEKVOA-IDEKIL-LIRAPFOYSDINAKYEENKEFNNOYALAKVEQILLOMOGY 96 |
| Db | 121 | QVLRFOD-EKA-LNVODKKRIFSVYLDLKGWGLIEDLAKNL--K-DEN--SPYLIDL-- 170 |
| QY | 97 | KVIVSDSDKDDDFEAKKKGGLYLAVANGELIYLRPDKRILOKSEGGLEFSTGLDMKG 150 |
| Db | 171 | VDOSGSGWENEFPEESNRVHDFEAVETGFOATYYTSTNNASGGFNSKSVIHNELD 230 |
| QY | 157 | VLIAGEIKYILIEPMGESIDSTMDLSELD-IOEKFLTHTSSHSGGLVSTWYKGT-D 214 |
| Db | 231 | KNREDAIHKILINRYAVVMKAVTELTKENIARYRDLMDMG 273 |
| QY | 215 | NSMDAIKSAIKNTFFANIOEIDKTLLOKNDLESTQDKAKELKG 256 |

(J.M.)

Distribution rights by Oxford Molecular Ltd

ch, using Smith-Waterman algorithm

MASPAT Time 23.59 seconds

Generalized

02 03 200-433-4
28-260) from US09308435 per

599

CSPH11EJNEVALKLNHPA.....QKNLESYQKDAKELKGRNR 233

ALM 150

11 ap

425878 seqs, 69334122 residues

Minimum Match Up

ptremb112

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

mean 46.340; Variance 102.801; scale 0.451

ed by analysis of the total score distribution.

SUMMARIES

| Entry | Accession | Length | DB | ID | Description | Pred. No. |
|-------|-----------|--------|--------|------------------------|-------------|-----------|
| 1 | 260 | 2 | 092L47 | NEURAMINYLACTOSE-BIND | 3.47e-19 | |
| 2 | 249 | 2 | 092L45 | POTATIVE PARALOG OF HP | 2.97e-12 | |
| 3 | 249 | 2 | 025L66 | POTATIVE NEURAMINYLAC | 5.35e-18 | |
| 4 | 282 | 2 | 092LX2 | POTATIVE PARALOG OF HP | 3.63e-10 | |
| 5 | 278 | 2 | 025334 | HYPOTHEICAL 31.9 KD P | 1.05e-08 | |
| 6 | 179 | 5 | 048453 | COMPLETE NUCLEOTIDE SE | 6.33e-05 | |
| 7 | 578 | 5 | 097281 | PC0695W PROTEIN. | 4.55e-02 | |
| 8 | 1027 | 5 | 006471 | COB. | 6.05e-02 | |
| 9 | 417 | 5 | 002080 | SMILAR TO CCAT/ENHAN | 1.07e-01 | |
| 10 | 549 | 14 | 071192 | 59 KDA PROTEIN. | 4.27e-01 | |
| 11 | 9376 | 2 | 085168 | SYRINGOMYCIN SYNTHETAS | 7.37e-01 | |
| 12 | 387 | 13 | 093382 | PAIRED TYPE HOMEOMAT | 1.38e-01 | |
| 13 | 944 | 2 | 047469 | FLACELLAR PROTEIN (FLB | 9.65e-01 | |
| 14 | 911 | 5 | 077338 | PC0565W PROTEIN. | 1.26e+00 | |
| 15 | 272 | 5 | 094895 | YOLK PROTEIN 1 (FRAGME | 1.65e+00 | |
| 16 | 402 | 1 | 058526 | GDP-BINDING PROTEIN HP | 1.65e+00 | |
| 17 | 501 | 10 | 064949 | ENDO-1,4-BETA GLUCANAS | 1.65e+00 | |
| 18 | 512 | 2 | 005176 | SUGAR TRANSFERER. | 1.65e+00 | |
| 19 | 394 | 2 | 044919 | FTSA. | 2.15e+00 | |
| 20 | 769 | 3 | 074253 | CELLOBIOSE DEHYDROGENA | 2.15e+00 | |

| | | | | | | |
|----|-----|-----|--------|--------|-------------------------|----------|
| 21 | 100 | 221 | 2 | 084553 | L3 RIBOSOMAL PROTEIN. | 4.71e+00 |
| 22 | 101 | 6.3 | 248 | 5 | YB18A.11 PROTEIN. | 3.6e+00 |
| 23 | 101 | 6.3 | 319 | 2 | FUNCTION UNKNOWN. | 3.65e+00 |
| 24 | 100 | 6.3 | 399 | 12 | AR15AL558 HOMEBOX-4 | 3.65e+00 |
| 25 | 101 | 6.3 | 383 | 13 | SODIUM-CALCIUM EXCHANG | 4.71e+00 |
| 26 | 101 | 6.3 | 592 | 1 | ADRI1850 | 3.65e+00 |
| 27 | 101 | 6.3 | 026557 | 9 | SENSOR TRANSDUCTION R | 3.65e+00 |
| 28 | 101 | 6.3 | 945 | 5 | POTASSIUM VOLTAGE-AND C | 3.65e+00 |
| 29 | 100 | 6.3 | 1803 | 3 | CHROMOSOME X READING F | 4.71e+00 |
| 30 | 99 | 6.2 | 202 | 2 | SEVING CYCLE ENZYME, P | 6.09e+00 |
| 31 | 99 | 6.2 | 235 | 2 | INVASIN. | 6.09e+00 |
| 32 | 99 | 6.2 | 559 | 2 | TRANSPOSASE TNIA. | 6.09e+00 |
| 33 | 99 | 6.2 | 559 | 12 | TRANSPOSASE OF IN21. | 6.09e+00 |
| 34 | 99 | 6.2 | 571 | 2 | TRANSPOSASE OF IN21. | 6.09e+00 |
| 35 | 99 | 6.2 | 571 | 2 | TRANSPOSASE OF IN21. | 6.09e+00 |
| 36 | 98 | 6.1 | 202 | 2 | YNGO PROTEIN. | 6.09e+00 |
| 37 | 98 | 6.1 | 269 | 2 | OOTER SURFACE PROTEIN | 7.87e+00 |
| 38 | 97 | 6.1 | 277 | 5 | TRANSRIBITION REGULA | 7.87e+00 |
| 39 | 97 | 6.1 | 405 | 10 | YOLK PROTEIN 1 (FRAGME | 1.01e+01 |
| 40 | 98 | 6.1 | 505 | 5 | PATATIN-LIKE PROTEIN (| 1.01e+01 |
| 41 | 97 | 6.1 | 559 | 2 | PF6A8.2 PROTEIN. | 1.01e+01 |
| 42 | 97 | 6.1 | 877 | 3 | FNIA PROTEIN. | 7.87e+00 |
| 43 | 98 | 6.1 | 877 | 3 | R32A7.6 PROTEIN. | 1.01e+01 |
| 44 | 97 | 6.1 | 980 | 3 | HP07HEM1 98.3 KD P | 1.01e+01 |
| 45 | 97 | 6.1 | 1312 | 4 | DEVELOPMENTAL REGULA | 7.87e+00 |
| 46 | 98 | 6.1 | 1790 | 3 | ANAXIN-2. | 1.01e+01 |
| 47 | 98 | 6.1 | 1790 | 3 | PH1/PH2 PROTEIN HOMOL | 7.87e+00 |

ALIGNMENTS

| ID | 1 | PRELIMINARY; | PRT; | 260 AA. |
|-----------------------|---|----------------------|---------------|-------------------|
| AC | 092L47; | | | |
| DT | 01-MAY-1999 (TReMBLrel..10, Created) | | | |
| DT | 01-MAY-1999 (TReMBLrel..10, Last sequence update) | | | |
| DT | 01-MAY-1999 (TReMBLrel..10, Last annotation update) | | | |
| DE | HEMAGGLUTININ-BINDING HEMAGGLUTININ PRECURSOR. | | | |
| GN | HPA. | | | |
| OS | Helicobacter pylori J99. | | | |
| OC | Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; | | | |
| CC | Helicobacter. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-J99; | | | |
| RX | MEDLINE; 99120557. | | | |
| RA | ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C., | | | |
| RA | SMITH D.R., NOONAN B., GOLD B.C., DEJONGE B.L., CARMEL G., | | | |
| RA | TOMLIN P.J., CARDO A., URNA-NICKELSEN M., MILLS D.M., IVES C., | | | |
| RA | GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOYIS G.F., | | | |
| RA | TRIST T.J.; | | | |
| RT | "Genomic-sequence comparison of two unrelated isolates of the human | | | |
| RT | gastric pathogen Helicobacter pylori."; | | | |
| RL | Nature 397:176-180(1999). | | | |
| DR | EMBL; AE001504; AAD06306.1; - | | | |
| SQ | SEQUENCE 260 AA; 29199 MW; 9C0103A4 CRC32; | | | |
| Query Match | 97.58; | Score 1559; | DB 2; | Length 260; |
| Best Local Similarity | 96.18; | Pred. No. 3,46e-262; | | |
| Matches 224; | Conservative | 7; | Mismatches 2; | Indels 0; Gaps 0; |
| Db | 28 CSPHIITETNVALKINWHPASEKVOALDEITLLLRPAFOYSDINAEYENKERNQTTLYK 87 | | | |
| QY | 28 CSHIITETNVALKINWHPASEKVOALDEITLLLRPAFOYSDINAEYENKERNQTTLYK 87 | | | |
| Db | 88 EELIIONGCVIVINWSSDKDDEFSAQKKEGYLAVANNGEIVLPPDKRTIÖKKSEPGILF 147 | | | |
| QY | 88 EQLIÖNGCVIVISVDSDDPFSFAQKKEGYLAVANNGEIVLPPDKRTIÖKKSEPGILF 147 | | | |
| Db | 148 STGLDMEGYVLIPAGVVKVYVILTEPMGESLDFETMDLSEHDIOEKFLKTTSHSGGLVS 207 | | | |
| QY | 148 STGLDMKEGYVLIPAGIKVYILTEPMGESLDFETMDLSEHDIOEKFLKTTSHSGGLVS 207 | | | |
| Db | 208 TWVKGNDNSDAIKSLNKFASINQEMDKLTÖRULSTESYÖKDAKLKKRN 260 | | | |


```

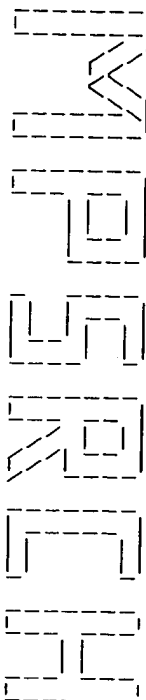
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
(1)
SEQUENCE FROM N.A.
MEDLINE: 96120851.
KAMBYSELLIS M.P., HO K.F., CRADDOCK E.M., PIANO F., PARISI M.,
COHEN J.,
"Pattern of ecological shifts in the diversification of Hawaiian
Drosophila inferred from a molecular phylogeny."
Curr. Biol. 5:1129-1139(1995).
[2]
SEQUENCE FROM N.A.
HO K.F.;
Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
EMBL: U51875; AAC47246.1;
FLYBASE: FBgn0016257; Dm1cYp1.
PFAM: PF00151; 1lpase; 1.
NON_TER 1 272
SEQUENCE 272 AA; 30187 MW; DF3EFC7C CRC32;

Query Match 6.5%; Score 104; DB 5; Length 272;
Best Local Similarity 29.1%; Pred.No.1.65e+00;
Matches 16; Conservative 15; Mismatches 23; Indels 1; Gaps 1;

116 VTIVTGLPKTSEVKKANSKLVNAYMQRVNLKQQOQOGOKYKNDSE-ENNROR 169
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
206 VSTVWKGTGDNNDIKRALNKIFANIMGEIDKRLQKNLESTQKDAELKGRNR 260

Arch completed: Tue Sep 26 16:20:56 2000
b time : 25 secs.

```



(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 26 16:16:34 2000; Maspar time 23.61 seconds

Tabular output not generated. 684.174 Million cell updates/sec

Title: >US-09-308-435-2

Description: (28-260) from US09308435.pep

Perfect Score: 1594

Sequence: 1 CSPHIETNEVALKLNHPA.....QKNLESYQKADKELGKRR 233

Scoring table: PAM 150

Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl12
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 46.291; Variance 102.838; scale 0.450

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-------------------------------|-----------|
| 1 | 1564 | 98.1 | 260 | 2 | Q9ZL47 NEURAMINYLACTOSE-BIND | 8.59e-263 |
| 2 | 230 | 14.4 | 249 | 2 | Q9ZK65 PUTATIVE PARALOG OF HP | 1.92e-18 |
| 3 | 222 | 13.9 | 249 | 2 | Q2S166 PUTATIVE NEURAMINYLACT | 3.51e-17 |
| 4 | 171 | 10.7 | 282 | 2 | Q9ZL47 PUTATIVE PARALOG OF HP | 1.99e-09 |
| 5 | 161 | 10.1 | 278 | 2 | Q2S234 HYPOTHETICAL 31.9 KD P | 5.54e-08 |
| 6 | 134 | 8.4 | 179 | 9 | Q48453 COMPLETE NUCLEOTIDE SE | 4.54e-02 |
| 7 | 117 | 7.3 | 578 | 5 | Q97281 PFC0895W PROTEIN. | 4.25e-01 |
| 8 | 109 | 6.8 | 549 | 14 | Q71192 59 KDA PROTEIN. | 4.25e-01 |
| 9 | 109 | 6.8 | 911 | 5 | Q71338 PFC0565W PROTEIN. | 4.25e-01 |
| 10 | 105 | 6.6 | 248 | 5 | Q9XW20 FLAGELLAR PROTEIN (FLB | 1.26e+00 |
| 11 | 105 | 6.6 | 394 | 2 | Q44769 FLAGELLAR PROTEIN (FLB | 9.60e-01 |
| 12 | 106 | 6.6 | 312 | 2 | Q05176 SUGAR TRANSPORTER. | 9.60e-01 |
| 13 | 106 | 6.6 | 1027 | 2 | Q06471 CROB. | 1.26e+00 |
| 14 | 105 | 6.6 | 1803 | 3 | P87192 CHROMOSOME X READING F | 2.14e+00 |
| 15 | 103 | 6.5 | 394 | 2 | Q44919 FTSA. | 2.14e+00 |
| 16 | 104 | 6.5 | 417 | 5 | Q02080 SIMILAR TO CCAAT/ENHAN | 1.64e+00 |
| 17 | 104 | 6.5 | 501 | 10 | Q64949 ENDO-1,4-BETA GLUCANAS | 1.64e+00 |
| 18 | 103 | 6.4 | 769 | 3 | Q74253 CELLOBIOSE DEHYDROGENA | 2.14e+00 |
| 19 | 102 | 6.4 | 202 | 2 | Q08230 OUTER SURFACE PROTEIN | 2.78e+00 |
| 20 | 102 | 6.4 | 235 | 2 | Q47380 INVASIN. | 2.78e+00 |

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|----|-------------------------------|-----------|
| 21 | 102 | 6.4 | 559 | 12 | Q9WT11 TRANSPOSASE OF IN21. | 2.78e+00 |
| 22 | 102 | 6.4 | 559 | 2 | Q48382 TRANSPOSASE TNIA. | 2.78e+00 |
| 23 | 102 | 6.4 | 571 | 2 | Q56281 TNIA. | 2.78e+00 |
| 24 | 100 | 6.3 | 221 | 2 | Q84533 L3 RIBOSOMAL PROTEIN. | 4.68e+00 |
| 25 | 101 | 6.3 | 319 | 2 | Q45595 FUNCTION UNKNOWN. | 3.61e+00 |
| 26 | 101 | 6.3 | 559 | 2 | Q92E69 TNIA PROTEIN. | 3.61e+00 |
| 27 | 101 | 6.3 | 571 | 2 | Q32213 YVGO PROTEIN. | 3.61e+00 |
| 28 | 101 | 6.3 | 592 | 1 | Q26557 SENSORY TRANSDUCTION R | 3.61e+00 |
| 29 | 100 | 6.3 | 856 | 2 | Q9ZMH1 HEAT SHOCK PROTEIN. | 4.68e+00 |
| 30 | 101 | 6.3 | 945 | 5 | Q9Y1J9 PUTATIVE VOLTAGE-AND C | 3.61e+00 |
| 31 | 101 | 6.3 | 1000 | 9 | Q80181 PUTATIVE MINOR STRUCTU | 6.05e+00 |
| 32 | 99 | 6.2 | 272 | 5 | Q94895 YOLK PROTEIN 1 (FRAGME | 6.05e+00 |
| 33 | 99 | 6.2 | 405 | 10 | Q23784 PATATIN-LIKE PROTEIN (| 6.05e+00 |
| 34 | 97 | 6.1 | 604 | 5 | Q91863 P2247.6 PROTEIN. | 1.01e+01 |
| 35 | 98 | 6.1 | 845 | 10 | Q92V60 PUTATIVE DNA RECOMBINA | 7.82e+00 |
| 36 | 97 | 6.1 | 914 | 4 | Q99493 SCAL PROTEIN (FRAGMENT | 1.01e+01 |
| 37 | 98 | 6.1 | 980 | 3 | Q00760 DEVELOPMENTAL REGULATO | 7.82e+00 |
| 38 | 97 | 6.1 | 1312 | 4 | Q99700 ATRAXIN-2. | 1.01e+01 |
| 39 | 98 | 6.1 | 1790 | 3 | P78621 FHL/PH2 PROTEIN HOMOLO | 7.82e+00 |
| 40 | 97 | 6.1 | 9376 | 2 | Q85168 STRIGOMYCIN SYNTHETAS | 1.01e+01 |
| 41 | 96 | 6.0 | 110 | 8 | Q9XUY9 MATURASE K (FRAGMENT) | 1.30e+01 |
| 42 | 96 | 6.0 | 231 | 1 | Q58888 HYPOTHETICAL PROTEIN M | 1.30e+01 |
| 43 | 96 | 6.0 | 252 | 2 | Q49544 VAA SURFACE LIPOPROTEI | 1.30e+01 |
| 44 | 96 | 6.0 | 331 | 5 | Q21938 R1D1.6 PROTEIN. | 1.30e+01 |
| 45 | 96 | 6.0 | 768 | 3 | Q42729 CELLOBIOSE DEHYDROGENA | 1.30e+01 |

ALIGNMENTS

RESULT 1 PRELIMINARY: PRT: 260 AA.

ID Q9ZL47. 01-MAY-1999 (TREMBLrel. 10, Created)
AC Q9ZL47. 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE NEURAMINYLACTOSE-BINDING HEMAGGLUTININ PRECURSOR.
GN HPA.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J99.
RX MEDLINE: 99120557.

RA ALM R.A., LING L.-S., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUIDI B.C., DEJONGE B.L., CARELL G.,
RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA TRUST T.U.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL: AE001504; AAD06306.1;
SQ SEQUENCE 260 AA; 29199 MW; 9C0103A4 CRC32;

Query Match 98.1%; Score 1564; DB 2; Length 260;
Best Local Similarity 96.6%; Pred. No. 8.59e-263;
Matches 225; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

| DB | Score | Query Match | Length | ID | Description | Pred. No. |
|----|-------|--------------------|--------------------|-----------------------|-------------|-----------|
| DB | 28 | CSPHIETNEVALKLNHPA | SEKYOALDEKILLRPA | FQYSDNIAKEYENFKNOTLKV | 87 | |
| QY | 28 | CSPHIETNEVALKLNHPA | SEKYOALDEKILLRPA | FQYSDNIAKEYENFKNOTLKV | 87 | |
| DB | 88 | EELIONGYVIVNDSDKDD | ESFQKKEGGLAAMGEYLR | PRDPKRTIKKSEPGILF | 147 | |
| QY | 88 | EELIONGYVIVNDSDKDD | ESFQKKEGGLAAMGEYLR | PRDPKRTIKKSEPGILF | 147 | |
| DB | 148 | STGLDKMEGYLIPAGFYK | YVILPEPSEGLDSFTMD | SELDIOKFKTTHSSSGGLVS | 207 | |
| QY | 148 | STGLDKMEGYLIPAGFYK | YVILPEPSEGLDSFTMD | SELDIOKFKTTHSSSGGLVS | 207 | |
| DB | 208 | TMYKGDNSDAIKSALNKI | TFASIMQEMDKLTO | RNLESYQKADKELGKRR | 260 | |

102 DSSDDDFEFAQKKEEYLA VAMNGEYLARDPKRTIQKKEPILFSTGLDKMEGLIPA 161
503 YLTTLNLHDKAFTRRNLDGNKDKGFSDLRIEENFLKSA 540
162 GFIKVTIL-E-PMSEGLSDFTMDLSELDIQEFKLT 197
[1] 01-NOV-1998 (TREMblrel. 08, Created)
01-NOV-1998 (TREMblrel. 08, Last sequence update)
01-NOV-1999 (TREMblrel. 12, Last annotation update)
PFC0565W PROTEIN.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
[1] SEQUENCE FROM N.A.
STRAIN-3D7;
HAMILIN N., LAWSON D., BARRELL B.;
Submitted (OCT-1998) to the EMBL/genbank/DBJ databases.
EMBL, AL008970, CA15593.1;
SEQUENCE 911 AA; 108563 MW; 5043B736 CRC32;
Query Match 6.8%; Score 109; DB 5; Length 911;
Best Local Similarity 19.8%; Pred. No. 4.25e+01;
atches 25; Conservative 44; Mismatches 52; Indels 5; Gaps 5;
703 NEDLIDENNRYTIRKSDICITKEIKNNNNINLNKVKIKFYLLKEKNNIIFVNR1 762
106 KDDSEFAQKKEEYLA VAMNGE-IVLRPDKRTIQ-KKSPGLFSTGLDKMEGLIPAGF 163
763 DLIITNEKKRNDLQES-NLFN-DIPITLTKNTNTINLKIITIHKNNTYTST 820
164 -IKVITLPMGESLSDFTMDLSELDIQEFKLTGSHSGGLVSTWVGTDNSDAIKS 222
821 SLNLF 826
223 ALNRIE 228
ULT 10
O9XM20 PRELIMINARY; PRT; 248 AA.
O9XM20;
01-NOV-1999 (TREMblrel. 12, Created)
01-NOV-1999 (TREMblrel. 12, Last sequence update)
01-NOV-1999 (TREMblrel. 12, Last annotation update)
Y18D10A.11 PROTEIN.
Y18D10A.11.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiia; Rhabditiida; Rhabditiida; Peloderiinae; Caenorhabditiis.
[1] SEQUENCE FROM N.A.
MEDLINE: 94150718.
HARRIS B.;
"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
Submitted (DEC-1998) to the EMBL/genbank/DBJ databases.
[2] SEQUENCE FROM N.A.
MEDLINE: 94150718.
WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERTS M.,
BONFIELD J., BUTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
GARDNER A., GREEN P., HARKINS T., HILLIER L., JIER M., JOHNSTON L.,
JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LARRELL P.,
LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
PARSONS J., PERCY C., RIFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
SMALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,
THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RT Nature 368:32-38(1994).
DR EMBL: AL034393; CA122314.1;
SQ SEQUENCE 248 AA; 27963 MW; BC3B5F9C CRC32;
Query Match 6.6%; Score 105; DB 5; Length 248;
Best Local Similarity 22.0%; Pred. No. 1.26e+00;
Matches 28; Conservative 40; Mismatches 51; Indels 8; Gaps 8;
Db 74 ORRWEGELASSRAQNEQL-SEKSOLOKENE-ELLVL-L-RTGIVDANKSTSEQLANA 129
113 ORK-BEGLAVAMNGEIVLRPDKRTIQKKEPILFSTGLDKMEGLIPAGFIKVTILP 171
QY 130 QDAEVSMTLKTETVAKLEENNVLTQATMAERSGLVNELIDTKELAKSAQOTELDESH 189
QY 172 MSGE-SLDSFTMDLSELDIQEFKLTGSHSGGLVSTWVGTDN-SNDA-1KSALENKIF 228
Db 190 RDMTAV 196
QY 229 ANIMOEI 235
RESULT 11
ID 044769 PRELIMINARY; PRT; 394 AA.
AC 044769;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE FLAGELLAR PROTEIN (FLBC) (ORF20).
GN BB0285
OC Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
[1] SEQUENCE FROM N.A.
RP MEDLINE: 98065943.
RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGR A., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN-VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTERBACK T., MATTHEY L., MCDONALD L., ARTTACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi."
RT Nature 390:580-586(1997).
[2] SEQUENCE FROM N.A.
RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGR A., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN-VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTERBACK T., MATTHEY L., MCDONALD L., ARTTACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
[3] SEQUENCE OF 3-394 FROM N.A.
RC STRAIN-B31;
RA DURN J.J., BUTLER-LOFREDO L., KIELECAZAWA J., MEDALLE J., LUFT B.J.;
RL Submitted (DEC-1995) to the EMBL/genbank/DBJ databases.
DR EMBL: AE001137; AAC66663.1;
DR EMBL: U43739; AAA85608.1;
DR HSSP: P13466; 1KSR.
DR TIGR: BB0285;
KW FLAGELLA.
SQ SEQUENCE 394 AA; 45073 MW; 5BF8FA60 CRC32;
Query Match 6.6%; Score 106; DB 2; Length 394;
Best Local Similarity 19.7%; Pred. No. 9.60e-01;

SEQUENCE 394 AA; 45005 MW; 57F04C42 CRC32;

```

Query Match      6.58; Score 103; DB 2; Length 394;
Post Total 441414

```

```

test Local Similarity 20.9%; Pred.No. 2.14e+00;
atches 24; Conservative 40; Mismatches 45; Indels 6; Gaps 6;

```

| | | | | | | | | | |
|---------|-----|--------------|-----|------------|-----|--------|----|------|----|
| Matches | 24; | Conservative | 40; | Mismatches | 45; | Indels | 6; | Gaps | 6; |
|---------|-----|--------------|-----|------------|-----|--------|----|------|----|

249 DFVSSLKPEWNLKINKNIYDKA-KVY-LKSNTGEIKLYLKPKELGSIRINLNLDSNNL 306


74 EYENKEFNQALKEQILNÖGKYIVSDSSDKDDEFAÖK-KE-GILAVAMNGEIVLRP 131

307 LGIIVDQNVK-MLPDQNMHSLNKMGGSGF-NASLNFLAGENLNSFTGDEKD 359

132 DPKRTIQKSEPGLLFSTGLDKMEGVLPAGFIKVTILEPMGSESLDSFTMDLSE 186

Arch. completed: Tue Sep 26 16:17:02 2000
 > time : 28 secs.

time : 28 secs.



 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

.Psrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Sep 26 16:18:57 2000; Maspar time 8.99 Seconds
 Tabular output not generated. 613.770 Million cell updates/sec

Title: >US-09-308-435-4
 Description: (28-260) from US09308435.pep
 Perfect Score: 1599
 Sequence: 1 CSPHIETNEVALKLNYPHA.....QKNLESYQDAKELKGRNR 233

Scoring table: PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseq36
 1:geneseqp

Statistics: Mean 32.470; Variance 175.985; scale 0.185

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description | Pred. No. |
|------------|-------|-------------|--------|--------|-------------------------|-----------|
| 1 | 1599 | 100.0 | 260 | W07450 | Helicobacter pylori su | 2.57e-113 |
| 2 | 1599 | 100.0 | 260 | W60157 | Helicobacter pylori 29 | 2.57e-113 |
| 3 | 1589 | 99.4 | 260 | W60156 | Helicobacter pylori 29 | 1.58e-112 |
| 4 | 1589 | 99.4 | 260 | W07449 | Helicobacter pylori su | 1.58e-112 |
| 5 | 1589 | 99.4 | 461 | W74466 | Adhesin/V. cholerae tox | 1.58e-112 |
| 6 | 1547 | 96.7 | 260 | Y11000 | H. pylori ORF 11a2071 | 3.20e-109 |
| 7 | 1540 | 96.3 | 268 | W20973 | H. pylori derived prot | 1.14e-108 |
| 8 | 302 | 18.9 | 97 | W24663 | H. pylori derived prot | 5.74e-13 |
| 9 | 302 | 18.9 | 97 | W20454 | H. pylori derived prot | 5.74e-13 |
| 10 | 235 | 14.7 | 249 | W55445 | H. pylori ORF 01a61101 | 3.94e-08 |
| 11 | 235 | 14.7 | 249 | W55332 | H. pylori ORF hp361107 | 3.94e-08 |
| 12 | 227 | 14.2 | 249 | W98332 | H. pylori ORF 1615 pr | 1.46e-07 |
| 13 | 176 | 11.0 | 282 | Y10951 | H. pylori ORF 04e61110 | 5.19e-04 |
| 14 | 176 | 11.0 | 285 | W20938 | H. pylori secreted or | 3.41e+00 |
| 15 | 118 | 7.4 | 147 | W20468 | H. pylori secreted or | 3.41e+00 |
| 16 | 109 | 6.8 | 549 | W21638 | Grapevine leafroll vir | 1.63e+01 |
| 17 | 107 | 6.7 | 615 | R77276 | ORC3 subunit of Yeast | 1.63e+01 |
| 18 | 107 | 6.7 | 615 | W22226 | S. cerevisiae oridin o | 1.63e+01 |
| 19 | 103 | 6.4 | 2039 | W56322 | Haemophilus paragailln | 2.86e+01 |
| 20 | 103 | 6.4 | 2042 | W56319 | Haemophilus paragailln | 2.86e+01 |
| 21 | 100 | 6.3 | 1254 | R07503 | Merizote apical-end-p | 4.33e+01 |
| 22 | 100 | 6.3 | 1254 | W24575 | Merizote apical-end p | 4.33e+01 |
| 23 | 99 | 6.2 | 1285 | P93089 | Pasteurella multocida | 4.97e+01 |

| | | | | | | |
|----|----|-----|------|---|--------|-------------------------|
| 24 | 97 | 6.1 | 119 | 1 | W49813 | Amino acid sequence of |
| 25 | 97 | 6.1 | 222 | 1 | R39267 | Humanised C4G1 Ig heav |
| 26 | 97 | 6.1 | 222 | 1 | W49817 | Fragment of humanised |
| 27 | 97 | 6.1 | 235 | 1 | R39268 | Humanised C4G1 Ig heav |
| 28 | 97 | 6.1 | 235 | 1 | W49818 | Human acid sequence of |
| 29 | 97 | 6.1 | 449 | 1 | R43339 | Human acid sequence of |
| 30 | 97 | 6.1 | 449 | 1 | W49816 | Human acid sequence of |
| 31 | 97 | 6.1 | 552 | 1 | W90172 | Human heart muscle spe |
| 32 | 97 | 6.1 | 552 | 1 | W24800 | Human heart muscle spe |
| 33 | 97 | 6.1 | 1312 | 1 | W3807 | Spinocerebellar ataxia |
| 34 | 97 | 6.1 | 1313 | 1 | W60213 | Human ataxin-2. |
| 35 | 96 | 6.0 | 246 | 1 | W95018 | Spinocerebellar ataxia |
| 36 | 96 | 6.0 | 947 | 1 | R76708 | Pentunia MADS box trans |
| 37 | 96 | 6.0 | 947 | 1 | R90924 | Recombinant cold-resis |
| 38 | 94 | 5.9 | 175 | 1 | R62783 | Corn derived pyruvate |
| 39 | 95 | 5.9 | 759 | 1 | W73479 | Borrelia VS461 antigen |
| 40 | 95 | 5.9 | 751 | 1 | W13491 | Grapevine leafroll vir |
| 41 | 95 | 5.9 | 764 | 1 | W98549 | Helicobacter CIPB (N-t |
| 42 | 94 | 5.9 | 1084 | 1 | R71015 | H. pylori GHPO 1604 pr |
| 43 | 94 | 5.9 | 1091 | 1 | R71015 | Human neuronal calcium |
| 44 | 95 | 5.9 | 1140 | 1 | R72386 | Human neuronal calcium |
| 45 | 94 | 5.9 | 1392 | 1 | Y06999 | XAP-1, part of the DNA |
| | | | | | | Reslin protein sequenc |

ALIGNMENTS

RESULT 1
 W07450 standard; Protein: 260 AA.
 AC W07450;
 DT 04-MAR-1997 (first entry)
 DE Helicobacter pylori surface-exposed 29 kDa antigen.
 KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
 therapy; diagnosis.
 OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
 FH Key
 Location/Qualifiers
 FT peptide
 1..27
 /label= Sig-peptide
 FT cleavage_site
 25..28
 /note="consensus recognition site for signal
 peptidase II in proilipoproteins"
 FT protein
 28..260
 /label= Mat.protein
 FT /note="either the uncleaved or mature protein can
 be used in appls. of the invention"

W09638475-A1.
 PD 05-DEC-1996.
 PF 03-JUN-1996; SE0727.
 PR 01-JUN-1995; SE-002007.
 PR 21-MAR-1996; SE-001085.
 PA (ASTR) ASTRA AB.
 PI Bolin I, Svennerholm A;
 DR N-Peds; T43835.
 DT WPI: 97-034307/03.
 PT Helicobacter pylori surface exposed antigen - useful for treatment,
 prophylaxis or diagnosis of mammalian infections
 PS Claim 2; Page 37-38; 49pp; English.
 CC A 29 kDa antigen (W07450) of Helicobacter pylori is an adhesin
 expressed on the surface of resting and dividing forms of the
 bacterium. It is conserved in all tested strains of H. pylori,
 CC and gives rise to both systemic and local (mucosal) prodn. of
 CC antibodies. Its amino acid sequence was deduced from a cDNA clone
 CC (T43835); a sequence (W07349) deduced from a second cDNA clone
 CC (T43834) differs only at position 222 (Ser for Arg). Recombinant
 CC antigen can be produced in transformed host cells. It is useful
 CC for the diagnosis and treatment of H. pylori infections and for
 CC use as a vaccine.
 SQ Sequence 260 AA;

Query Match 100.0%; Score 1599; DB 1; Length 260;
 Best Local Similarity 100.0%; Pred. No. 2.57e-113;
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 CSPHIETNEVALKLNYPASEKVALDEKILLRPAFOYSDNIKEYNKFNQTAIKV 87

```

QY 28 CSPHIIETNEVALKLNHPASEKVOALDEKILLRPAFOYSDNIAKEYENKFNQATLV 87
DB 88 EQLIIONOGYKIVSYDSSDKDPSFAOKKEGYLAVMNGEIVLRDPKRTIOKKEPGLLF 147
QY 88 EQLIIONOGYKIVSYDSSDKDPSFAOKKEGYLAVMNGEIVLRDPKRTIOKKEPGLLF 147
DB 148 STGLDKMEGVLPAGFIKVTILPEPMSGESLDSFTMDLSELDIOEKFLKTHSSHGGLVS 207
QY 148 STGLDKMEGVLPAGFIKVTILPEPMSGESLDSFTMDLSELDIOEKFLKTHSSHGGLVS 207
DB 208 TMVKGTDNSDAIKRALNKIFANIMOIEIDKKLTOKNLESYQDAKEKGRNR 260
QY 208 TMVKGTDNSDAIKRALNKIFANIMOIEIDKKLTOKNLESYQDAKEKGRNR 260

RESULT 2
ID W60157 standard; Protein; 260 AA.
AC W60157;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
NS MO9822135-A1.
PF 28-MAY-1998; SE1927.
PR 18-NOV-1997; SE1927.
PA (ASTR ) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34643.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Disclosure; Page 35-36; Sipp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (baccillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34643)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second, claimed 29 kDa antigen (see W60156) is
CC provided that differs only at residue 222 (Ser for Arg).
J Sequence 260 AA;

Query Match 100.0%; Score 1599; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.57e-113;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

ID W60156 standard; Protein; 260 AA.
AC W60156;
DT 28-SEP-1998 (first entry)
DE Helicobacter pylori 29 kDa surface-exposed antigen.
KW Cell surface antigen; adhesin; infection; vaccine; drug delivery.
NS MO9822135-A1.
PF 28-MAY-1998; SE1927.
PR 18-NOV-1997; SE1927.
PA (ASTR ) ASTRA AB.
PI Berglindh T, Lofroth J;
DR WPI; 98-312174/27.
DR N-PSDB; V34642.
PT Aggregate of negatively charged lipid and Helicobacter pylori
PT antigen - useful in therapeutic or preventative vaccines against
PT both forms of bacterium, to generate systemic and mucosal antibody
PT responses
PS Claim 6; Page 28-30; Sipp; English.
CC This polypeptide comprises a 29 kDa surface antigen which is
CC expressed on the surface of both dividing (baccillary) and resting
CC (colloid) forms of Helicobacter pylori. The antigen gives rise to
CC both systemic and local (mucosal) production of antibodies. It is
CC an adhesin that is conserved in all tested strains of H. pylori,
CC and is also a putative virulence factor. The invention relates to
CC a new pharmaceutical composition comprising a lipid aggregate of a
CC negatively charged lipid or lipid mixture and at least one
CC antigenic, native or recombinant antigen of H. pylori, preferably
CC the surface-exposed 29 kDa antigen, or a nucleic acid (see V34642)
CC coding for such an antigen. The new formulations, which are useful
CC as vaccine formulations, elicit a protective immune response
CC against H. pylori infections, and are suitable for therapeutic and
CC prophylactic use. An adequate response is produced even without
CC adjuvant. A second 29 kDa antigen (see W60157) is provided that
CC differs only at residue 222 (Arg for Ser).
SQ Sequence 260 AA;

Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1.58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 28 CSPHIIETNEVALKLNHPASEKVOALDEKILLRPAFOYSDNIAKEYENKFNQATLV 87
QY 28 CSPHIIETNEVALKLNHPASEKVOALDEKILLRPAFOYSDNIAKEYENKFNQATLV 87
DB 88 EQLIIONOGYKIVSYDSSDKDPSFAOKKEGYLAVMNGEIVLRDPKRTIOKKEPGLLF 147
QY 88 EQLIIONOGYKIVSYDSSDKDPSFAOKKEGYLAVMNGEIVLRDPKRTIOKKEPGLLF 147
DB 148 STGLDKMEGVLPAGFIKVTILPEPMSGESLDSFTMDLSELDIOEKFLKTHSSHGGLVS 207
QY 148 STGLDKMEGVLPAGFIKVTILPEPMSGESLDSFTMDLSELDIOEKFLKTHSSHGGLVS 207
DB 208 TMVKGTDNSDAIKRALNKIFANIMOIEIDKKLTOKNLESYQDAKEKGRNR 260
QY 208 TMVKGTDNSDAIKRALNKIFANIMOIEIDKKLTOKNLESYQDAKEKGRNR 260

RESULT 4
ID W07449 standard; Protein; 260 AA.
AC W07449;
DT 04-MAR-1997 (first entry)
DE Helicobacter pylori surface-exposed 29 kDa antigen.
KW Antigen; adhesin; immunogen; vaccine; immunisation; ulcer;
KW therapy; diagnosis.
OS Helicobacter pylori strain CCUG 17874 (NTCC 11637).
FH Key
FT Peptide 1..27
FT Location/Qualifiers
FT cleavage_site 25..28

```

FT /note- "consensus recognition site for signal
FT peptidase II in prolipoproteins"
FT 28. 260
FT /label- Mat.protein
FT /note- "either the uncleaved or mature protein can
FT be used in aplns. of the invention"
PD W09638475-A1.
PD 05-DEC-1996.
PR 03-JUN-1996; SE0727.
PR 01-JUN-1995; SE-002007.
PR 21-MAR-1996; SE-001085.
PA (ASTRA) ASTRA AB.
PI Bohn I, Svennerholm A;
DR WPI: 97-034307/03.
DR N-PSDB: T43834.
CC Helicobacter pylori surface exposed antigen - useful for treatment,
CC prophylaxis or diagnosis of mammalian infections
CC Claim 2; Page 34-35; 49pp; English.
CC A 29 kDa antigen (W07445) of Helicobacter pylori is an adhesin
CC expressed on the surface of resting and dividing forms of the
CC bacterium. It is conserved in all tested strains of H. pylori,
CC and gives rise to both systemic and local (mucosal) prodn. of
CC antibodies. Its amino acid sequence was deduced from a cDNA clone
CC (T43834); a sequence (W07550) deduced from a second cDNA clone
CC (T43835) differs only at position 222 (Arg for Ser). Recombinant
CC antigen can be produced in transformed host cells. It is useful
CC for the diagnosis and treatment of H. pylori infections and for
CC use as a vaccine.
SQ Sequence 260 AA;

Query Match 99.4%; Score 1589; DB 1; Length 260;
Best Local Similarity 99.6%; Pred. No. 1,58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 28 CSPHIETNEVALKLTNYHPASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 87
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 88 EQLLNQGYKIVSVSDSDKDFEFQAQKKEGYLVANNGEIVLRPDPKRTIOKSEGLLF 147
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 88 EQLLNQGYKIVSVSDSDKDFEFQAQKKEGYLVANNGEIVLRPDPKRTIOKSEGLLF 147
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 148 STGLDMKEGVLLPAGFIKYITILEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 148 STGLDMKEGVLLPAGFIKYITILEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 208 TMYKGTDSNDAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 208 TMYKGTDSNDAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5
ID W74466 standard; Protein; 461 AA.
AC W74466;
DT 18-MAY-1999 (first entry)
DE Adhesin/V.cholerae toxin A2 and B subunit fusion protein.
KW Fusion gene; H. pylori infection; adhesin gene; V. cholerae toxin;
KW A2 subunit; B subunit; diagnosis; vaccine; gastritis; gastric ulcer;
KW duodenal ulcer.
OS Chimeric - Helicobacter pylori.
OS Chimeric - Vibrio cholerae.
PN W09853082-A1.
PD 26-NOV-1998.
PD 21-MAY-1997; 000091.
PR 21-MAY-1997; WO-KR0091.
PA (DAEW-) DAEWOONG PHARM CO LTD.
PI Kim Bo, Lee BK, Park SK, Yoon SW, Yu YH;
DR WPI: 98-434925/37.
DR N-PSDB: X21955.
CC Fusion gene consisting of the Helicobacter pylori adhesin gene
CC ligated to the A2 and B subunit gene of Vibrio cholerae toxin -
CC useful in the diagnosis and treatment of H. pylori
CC Claim 2; Page 20-21; 32pp; English.

CC This sequence represents the fusion protein of the invention, between the
CC H. pylori adhesin gene, and the V. cholerae toxin A2 and B subunit genes.
CC This protein is used in the diagnosis of and as a vaccine against
CC H. pylori, which is the main cause of gastritis and gastric/duodenal
CC ulcers. It may also be used in the production of an anti-H. pylori
CC antibody. Vectors containing the DNA sequence and E. coli strains
CC containing it form the basis of a process to produce the protein, which
CC consists of culturing the microorganism and obtaining the protein. The
CC protein is an effective vaccine against H. pylori due to its excellent
CC immunogenicity for H. pylori, stability within the stomach environment,
CC and its ability to penetrate through the stomach mucous membrane to
CC stimulate siga production.
CC Note: This sequence was indexed from W09853082, which is the first
CC major country equivalent to KR97059278.
SQ Sequence 461 AA;

Query Match 99.4%; Score 1589; DB 1; Length 461;
Best Local Similarity 99.6%; Pred. No. 1,58e-112;
Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 39 CSPHIETNEVALKLTNYHPASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 98
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 28 CSPHIETNEVALKLTNYHPASEKVOALDEKILLRPAFYSDNIAEYENKFNQALKY 87
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 99 EQLLNQGYKIVSVSDSDKDFEFQAQKKEGYLVANNGEIVLRPDPKRTIOKSEGLLF 158
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 88 EQLLNQGYKIVSVSDSDKDFEFQAQKKEGYLVANNGEIVLRPDPKRTIOKSEGLLF 147
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 159 STGLDMKEGVLLPAGFIKYITILEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 218
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 148 STGLDMKEGVLLPAGFIKYITILEPMSESLDFTMDLSELDIOEKFLKTHSHSGGLVS 207
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 219 TMYKGTDSNDAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 271
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 208 TMYKGTDSNDAIKRALNKIFANIMOEDIKLTQKNLESYQDAKELKGRNR 260
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 6
ID Y11000 standard; Protein; 260 AA.
AC Y11000;
DT 08-JUN-1999 (first entry)
DE H. pylori ORF 11a2p20714_4797137_f3_45 cell envelope protein.
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW secreted protein; cellular protein.
OS Helicobacter pylori.
PN W09618323-A1.
PD 07-MAY-1998.
PD 28-OCT-1997; U19575.
PR 14-JUL-1997; US-891928.
PR 28-OCT-1996; US-739150.
PR 06-DEC-1996; US-759739.
PA (ASTRA) ASTRA AB.
PI Alm RA, Smith D;
DR WPI: 98-271811/24.
DR N-PSDB: X30467.
CC Helicobacter pylori nucleic acids and proteins - used to develop
CC products for the detection, prevention and treatment of H. pylori
CC infections
CC Claims 27, 31; Page 207-208; 279pp; English.
CC Recombinant or substantially pure preparations of H. pylori polypeptides
CC are disclosed, together with the nucleic acids encoding them. In all,
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
CC secreted proteins or other cellular proteins. Vaccines containing the
CC nucleic acids or proteins are claimed, as are probes containing at least
CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC for treating or reducing the risk of H. pylori infections, and the
CC probes can be used diagnostically for detecting the presence of
CC Helicobacter in a sample. The products are also of use in screening
CC for compounds having the ability to interfere with the H. pylori life
CC cycle or to inhibit H. pylori infection.
SQ Sequence 260 AA;

Query Match 96.7%; Score 1547; DB 1; Length 260;

CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions for
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SC Sequence 261 AA;

Query Match 14.7%; Score 235; DB 1; Length 261;
 Best Local Similarity 25.1%; Pred. No. 3,946-08;
 Matches 57; Conservative 75; Mismatches 73; Indels 22; Gaps 21;

Db 52 VELHRY-PIKQKQEPKNNHVLIDPKIEANKVIPENTQKEKSLFLQLSNFERKGY 110
 38 VALKINYPHASEKQVALDEK-ILLRPAFOYSDNIAKEYENKFKQTALKEQILQNGY 96
 111 SY-S-QFKVSELPDIEKALVLMNGNVAI-LE-D-IVEE-SDA--L-S-E-EKVID 159
 97 KYISVSSDKDDEFSQKKEGYLAVMNGEIVLRPDKRTIQKSPGLFSTGLDKMEG 156
 Db 160 -M-SGGYLNLFVEPKSEDIHSFGIDVSKIKAVIER-VEL-RTNSGGVPEKPTF 215
 157 VLIPAGFIKVTILPEMSESLDSTFMDLSEL-DIOEKFLKTHSSHSGGLVS-TM- 213
 Db 216 ETDHRAIKINNOAYHKVAVHTKELSKKMEHYEKVSEEMK-KRK 261
 214 DMSND-AIKRALNKIFANIMOEDIKLTQKNLESYQKAKELKGRN 259

RESULT 12
 ID W98332 standard; Protein; 249 AA.
 AC W98332;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPD 1615 protein.
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KM Peptic ulcer disease.
 OS Helicobacter pylori.
 PN W09843478-A1.
 PD 08-OCT-1998.
 PE 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 (HUMA-) HUMAN GENOME SCI INC.
 (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR WPI: 98-542293/46.
 DR N-PSDB: X14051.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8, Page 455-456, 2054pp; English.
 CC This sequence represents a Helicobacter pylori GHPD protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SC Sequence 249 AA;

Query Match 14.2%; Score 227; DB 1; Length 249;
 Best Local Similarity 25.0%; Pred. No. 1,466-07;
 Matches 58; Conservative 75; Mismatches 76; Indels 23; Gaps 22;

Db 35 NGESEVLEHRY-PIKQKQEPKNNHVLIEPRIEINKVIPESYQKEKSLFLQLSNFL 93
 34 EINE-VALKINYPHASEKQVALDEK-ILLRPAFOYSDNIAKEYENKFKQTALKEQIL 91
 Db 94 ERGQYGV-S-QFKVSELPDIEKALVLMNGNVAI-LE-D-IVEE-SDA--L-S-E- 142
 92 QNGGIVISVSSDKDDEFSQKKEGYLAVMNGEIVLRPDKRTIQKSPGLFSTGL 151
 Db 143 EKVID-M-SGGYLNLFVEPKSEDIHSFGIDVSKIKAVIER-VEL-RTNSGGVPEKPTF 198
 152 DNEGVILIRAGFIKVTILPEMSESLDSTFMDLSEL-DIOEKFLKTHSSHSGGLVS-TM 209
 Db 199 VHRKTDHDAIKRTIMNOAYHKVAVHTKELSKKMEHYEKVSEEMK-KRK 249
 210 V-KGTDSND-AIKRALNKIFANIMOEDIKLTQKNLESYQKAKELKGRN 259

RESULT 13
 ID Y10951 standard; Protein; 282 AA.
 AC Y10951;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF 04ee1108.3906963-fl-7 cell envelope protein.
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein.
 KM Secreted protein; cellular protein.
 OS Helicobacter pylori.
 PN W09818323-A1.
 PD 07-MAY-1998.
 PE 28-OCT-1997; U19575.
 PR 14-JUL-1997; US-891928.
 PR 28-OCT-1996; US-739150.
 PR 06-DEC-1996; US-759739.
 PA (ASPR) ASPRA AB.
 PI Alm RA, Smith D;
 DR WPI: 98-271811/24.
 DR N-PSDB: X30418.
 PT Helicobacter pylori nucleic acids and proteins - used to develop
 PT products for the detection, prevention and treatment of H. pylori
 PT infections
 PS Claims 27, 31; Page 151, 279pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
 CC secreted proteins or other cellular proteins. Vaccines containing the
 CC nucleic acids or proteins are claimed, as are probes containing at least
 CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful
 CC for treating or reducing the risk of H. pylori infections, and the
 CC probes can be used diagnostically for detecting the presence of
 CC Helicobacter in a sample. The products are also of use in screening
 CC for compounds having the ability to interfere with the H. pylori life
 CC cycle or to inhibit H. pylori infection.
 SC Sequence 282 AA;

Query Match 11.0%; Score 176; DB 1; Length 282;
 Best Local Similarity 25.1%; Pred. No. 5,196-04;
 Matches 56; Conservative 59; Mismatches 90; Indels 18; Gaps 14;

Db 64 PUDFNY-PI-HYQAPQNHVGIIMPRQVSDN-KPTIDFQDALINQITIEKRGY 120
 39 AKLNYHPASEKQVALDEK-ILLRPAFOYSDNIAKEYENKFKQTALKEQILQNGY 96
 Db 121 VYLRFOD-EKA-LNYQDKKKIPSVLDLKGWGLIDDKMNL--K-DPN--SPNIDL-- 170
 97 KYISVSSDKDDEFSQKKEGYLAVMNGEIVLRPDKRTIQKSPGLFSTGLDKMEG 156
 Db 171 VDQSGSVWFNFEPSNRVYHDFAVGTFOAITTYTSTNNAAGGFGNSKSVIHEND 230
 157 VLIPAGFIKVTILPEMSESLDSTFMDLSEL-DIOEKFLKTHSSHSGGLVSTWTKGT-D 214
 Db 231 KNRDAIHKILNRYAVYKAVTLEKENTAKYDAIDRKG 273
 215 NSNDAIKRALNKIFANIMOEDIKLTQKNLESYQKAKELKGRN 256

